

STATISTICAL LITERACY AMONG SECOND LANGUAGE ACQUISITION
GRADUATE STUDENTS

By

Talip Gonulal

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies—Doctor of Philosophy

2016

ABSTRACT

STATISTICAL LITERACY AMONG SECOND LANGUAGE ACQUISITION GRADUATE STUDENTS

By

Talip Gonulal

The use of statistics in second language acquisition (SLA) research has increased over the past 30-40 years (Brown, 2004; Loewen & Gass, 2009). Further, several methodological syntheses (e.g., Plonsky, 2011; Plonsky & Gonulal, 2015; Winke, 2014) revealed that researchers in the field have begun to use more sophisticated and novel statistical methods (e.g., factor analysis, mixed models/mixed regression analyses, structural equation modeling, Bayesian statistics) even if common inferential statistics (e.g., *t* tests, ANOVAs, and correlations) are still dominating quantitative second language research (Plonsky, 2013, 2015). However, the increased use of a larger variety of statistical methods does not necessarily translate to high methodological quality. In fact, several SLA researchers have accentuated the state of statistical literacy and statistical training in the field of SLA (e.g., Godfroid & Spino, 2015; Loewen et al., 2014; Norris, Ross & Schoonen, 2015; Plonsky, 2011, 2013, 2015, Plonsky & Gonulal, 2015). Indeed, statistical literacy appears to be critical to SLA researchers' ability to advance L2 theory and practice. While some studies on statistical literacy in the field have been published, it appears that no studies exist that measure SLA researchers' statistical knowledge, which is also an important piece of the puzzle.

In this dissertation, I focus on SLA doctoral students—an important part of academia—and attempt to investigate their statistical training and knowledge of statistics. To this end, I used two primary instruments: the *SLA for SLA* (that is, the

Statistical Literacy Assessment for Second Language Acquisition) survey, and semi-structured interviews). One hundred and twenty SLA doctoral students in North America took the *SLA for SLA* survey, and 16 of them participated in follow-up interviews. The participants were from 30 different SLA programs across North America.

The results of this study show that doctoral students are well trained in basic descriptive statistics, while their training in inferential statistics, particularly advanced statistics, is limited. Further, it appears that self-training in statistics is not very common among SLA doctoral students. The results also point out that more in-house statistics courses, particularly intermediate and advanced statistics, are needed. When looking at their statistical knowledge, the results indicate that SLA doctoral students are good at understanding descriptive and inferential statistics, but they find it hard to interpret statistical analyses related to inferential statistics that are commonly encountered in SLA research. Another important finding is that as might be expected, the number of statistics courses taken, self-training in statistics and quantitative research orientation are predictive of statistical literacy, whereas surprisingly years spent in the doctoral program are significant components of statistical literacy. Based on the findings of this study, I make some suggestions directed toward improving statistical literacy in the field of SLA.

To all statisticians!

ACKNOWLEDGMENTS

I would like to express my appreciation to several people for their support during my academic journey in a place far from my home. First of all, I am intellectually indebted to my current advisor and committee chair, Dr. Shawn Loewen for his valuable support and feedback. I would not imagine that the quantitative research methods course that I took with him in my first semester would have such a significant effect on shaping my academic research interests. I am also grateful to my former advisor, Dr. Paula Winke for her time and guidance. She has always been helpful, supportive, and encouraging. I would also like to thank Dr. Aline Godfroid who provided helpful feedback and suggestions on the dissertation proposal. I am grateful to have such a good statistician on my dissertation committee. My special thanks also go to Dr. Susan Gass whom I am very fortunate to have on my dissertation committee. I thank Dr. Gass for her precious time and feedback. My gratitude also goes to Dr. Luke Plonsky whose work on methodological quality has motivated me to focus on statistical literacy. Thank you Luke!

I am also very grateful to the Turkish Ministry of Education for supporting me financially during my graduate studies. Many thanks also go to my colleagues in the doctoral program for their support during my academic and social development, especially Ina Choi, Yaqiong Cui, Lorena Valmori, Ji-Hyun Park, and my table tennis partner, Hyung-Jo Yoon.

Finally, special thanks also go to Biggby and Espresso Royale baristas, without your coffee, this study would have not been completed. Teşekkür ederim!

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	x
KEY TO ABBREVIATIONS.....	xi
CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW	1
1.1 The Use of Statistics in SLA.....	3
1.2 Methodological Quality in SLA.....	6
1.3 Graduate Training in Quantitative Research.....	9
1.4 Statistical Literacy	14
1.4.1 Statistical literacy and other related terms	15
1.4.2 Research on statistical literacy.....	19
1.4.3 Statistical literacy in SLA	22
1.5 Research Questions.....	24
CHAPTER 2: METHOD	26
2.1 Participants.....	26
2.2 Instruments.....	29
2.2.1 Statistical background questionnaire	29
2.2.2. Development of a discipline-specific statistical literacy assessment.....	29
2.2.2.1 Statistical literacy assessment for second language acquisition survey....	31
2.2.2.2 Pilot test	35
2.2.3 Semi-structured interviews	37
2.3 Procedure	39
2.4 Quantitative Data Analysis	40
2.4.1 Descriptive statistics	40
2.4.2 Missing data analysis	40
2.4.2.1 Multiple imputation	44
2.4.3 Exploratory factor analysis	44
2.4.3.1 Factorability of the data	45
2.4.3.2 Factor extraction model	46
2.4.3.3 Factor retention criteria.....	47
2.4.3.4 Factor rotation method.....	47
2.4.3.5 Interpretation of factors.....	48
2.4.4 Multiple regression analysis	48
2.5 Qualitative Data Analysis	50
CHAPTER 3: RESULTS	51
3.1 Research Question 1	51
3.2 Research Question 2	56
3.3 Research Question 3	65

3.4 Research Question 4	73
3.4.1 Lack of deeper statistical knowledge	74
3.4.2 Limited number of discipline-specific statistics courses	76
3.4.3 Major challenges in using statistical methods	78
3.4.4 Mixed-methods research culture	82
CHAPTER 4: DISCUSSION.....	85
4.1 Statistical Training in SLA	85
4.2 Statistical Literacy in SLA.....	89
4.3 Predictors of Statistical Literacy.....	94
4.4 A Glimpse into Pandora’s Box: Issues Related to Statistical Training and Using Statistics	97
4.5 Limitations	102
4.6 Suggestions for the Field of SLA.....	103
4.6.1 Improve statistical training in SLA.....	104
4.6.2 Increase the number of SLA faculty specializing in statistics	105
4.6.3 Increase students’ awareness of quantitative methods for SLA	106
CHAPTER 5: CONCLUSION	108
NOTES.....	109
APPENDICES.....	111
APPENDIX A SLA and Applied Linguistics Programs.....	112
APPENDIX B Background Questionnaire	113
APPENDIX C The SLA for SLA Instrument.....	117
APPENDIX D Interview Questions.....	127
APPENDIX E Survey Invitation Email.....	128
APPENDIX F Interview Invitation Email	130
APPENDIX G Sample Worry Questions about Statistical Messages (Gal, 2002)....	131
REFERENCES	133

LIST OF TABLES

Table 1 Current statistics self-efficacy by Finney and Schraw (2003, p.183).....	32
Table 2 List of the content domains addressed in the SLA for SLA instrument.....	34
Table 3 Interviewee Data.....	38
Table 4 Multiple Regression Assumptions.....	49
Table 5 Descriptive statistics for research orientation.....	52
Table 6 Overall statistical training.....	54
Table 7 Type and frequency of statistical assistance.....	55
Table 8 Type of statistical computation.....	56
Table 9 Item analysis on the SLA for SLA survey.....	57
Table 10 Factor loadings.....	62
Table 11 Descriptive statistics for factors.....	64
Table 12 Regression model summary for Factor 1.....	66
Table 13 Model data for Factor 1.....	66
Table 14 Alternative regression model summary for Factor 1.....	67
Table 15 Alternative model data for Factor 1.....	67
Table 16 Regression model summary for Factor 2.....	68
Table 17 Model data for Factor 2.....	68
Table 18 Alternative regression model summary for Factor 2.....	68
Table 19 Alternative model data for Factor 2.....	69
Table 20 Regression model summary for Factor 3.....	69
Table 21 Model data for Factor 3.....	70

Table 22	Alternative regression model summary for Factor 3	70
Table 23	Alternative model data for Factor 3	71
Table 24	Regression model summary for overall score.....	71
Table 25	Model data for overall score	71
Table 26	Alternative regression model summary for overall score.....	72
Table 27	Alternative model data for overall score.....	73
Table 28	List of doctoral programs conferring degrees in SLA and applied linguistics	112
Table 29	The raw data for the consensus task	119
Table 30	Descriptive statistics for all three tasks.....	120
Table 31	The results of the multiple regression analysis.....	125

LIST OF FIGURES

Figure 1. Geographic information about the participants	28
Figure 2. Example item on the Statistics Concept Inventory (Allen, 2006, p. 433).	30
Figure 3. Example item on the Statistical Literacy Inventory (Schild, 2002, p. 2).....	30
Figure 4. Items analysis on the second version of SLA for SLA survey	37
Figure 5. Missing value analysis (MVA).....	42
Figure 6. Departments in which statistics courses were taken.....	52
Figure 7. Participants' research orientation	53
Figure 8. Scree plot for 6-component solution	59
Figure 9. Visual comparison of factor retention criteria.....	61
Figure 10. Map of the United States and Canada	114
Figure 11. Graphs for map task data.....	120
Figure 12. Boxplots for questions 9 and 10	121

KEY TO ABBREVIATIONS

AL	Applied linguistics
CI	Confidence intervals
EFA	Exploratory factor analysis
ELL	English language learner
EV	Eigenvalue
KMO	Kaiser-Meyer-Olkin measure of sampling adequacy
L2	Second language
M	Mean
MA	Master of arts
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random
MVA	Missing value analysis
PCA	Principal components analysis
PhD	Doctor of philosophy
QUAL	Qualitative
QUAN	Quantitative
SCI	Statistics concept inventory
SD	Standard deviation
SEM	Structural equation modeling
SLA	Second language acquisition

SLA for SLA	Statistical literacy assessment for second language acquisition
SLI	Statistical literacy inventory
SRA	Statistics reasoning assessment
TEFL	Teaching English as a foreign language
TESOL	Teaching English to speakers of other languages
VIF	Variance inflation factors

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Second language acquisition (SLA¹) is a relatively new, yet developing, field. Indeed, the foundation of the first doctoral program in SLA (i.e., Department of Second Language Studies at the University of Hawai'i) goes back to 1988 (Thomas, 2013). SLA largely draws from other disciplines, as any developing field does (Selinker & Laksmanan, 2001). Although the use of quantitative research methods has been prevailing from the beginning, the field has seen an exponential increase in the use of statistical procedures in the last two decades, which Plonsky (2015) called a “methodological and statistical reform movement” (p. 4). For example, the pace at which relatively new and sophisticated statistical methods (e.g., factor analysis, structural equation modeling, mixed regression models) are used in second language (L2²) research has noticeably increased (Plonsky, 2015; Plonsky & Gonulal, 2015; Winke, 2014). In addition, there is a growing number of article- and book-length sources (e.g., Larson-Hall, 2010, 2015; Mackey & Gass, 2015; Plonsky, 2015) dealing with discipline-specific statistics and quantitative research designs.

As the field of SLA grows and develops, researchers have begun to draw on more and more advanced statistical methods. In fact, a number of scholars have attended to the quality of statistical knowledge and methodology in the field (e.g., Godfroid & Spino, 2015; Larson-Hall & Plonsky, 2015; Loewen et al., 2014; Norris, 2015; Norris, Ross & Schoonen, 2015; Plonsky, 2011, 2013, 2015; Plonsky & Gonulal, 2015). Indeed, given the strong quantitative research tradition and importance of statistics in the field, statistical literacy is necessary for the future development of the field and therefore it is

important for both established researchers and the future professoriate in the field of SLA. To reliably and accurately inform L2 theory and practice, established and developing L2 researchers need to have the skills and knowledge necessary to (a) choose the correct statistical methods suitable for their research, (b) conduct the statistical analyses appropriately, (c) engage in transparent reporting practices, (d) comprehend the results of research, and (e) evaluate the soundness of statistical analyses (Gonulal, Loewen & Plonsky, in preparation).

In addition, a few SLA researchers (Plonsky, 2011, 2013; Plonsky & Gonulal, 2015; Norris, 2015; Norris et al., 2015) have, to some extent, attributed the current state of methodological and statistical quality in L2 research to the limited state of statistical literacy in the field. Further, several voices mostly in sister disciplines such as psychology and education have argued that the development of statistical literacy depends somewhat on the quality of the statistical training that researchers receive in graduate programs (Aiken, West & Millsap, 2008; Capraro & Thompson, 2008; Gonulal et al., in preparation; Henson, Hull & Williams, 2010). Given that, it is unfortunate that the field of SLA has seen little research investigating the statistical knowledge of L2 researchers. To my knowledge, only two studies (i.e., Lazaraton, Riggenbach & Ediger, 1987; Loewen et al., 2014) have focused on the statistical literacy among SLA professors and graduate students. Although these two studies surface as playing a pioneering role in the investigation into the state of statistical literacy in the field, the studies are limited in several ways. First, in both studies, the researchers relied on self-report instruments to collect data about the statistical knowledge of L2 researchers. However, researchers' ability to interpret and use statistical procedures might be different from what they

assume they can do: They might over or underestimate their ability. Therefore, to accurately measure statistical literacy, instruments that can provide direct evidence of participants' statistical capabilities should be used. Second, because the researchers of both studies attempted to provide a broad picture of statistical literacy among L2 researchers, they included samples from two different populations: professors and graduate students. However, considering the potentially different experiences of professors and graduate students in using statistical procedures, it can be assumed that the statistical literacy level of these two groups *would* be different. While the question of SLA faculty's experience with quantitative research methods is a worthy area of investigation, an investigation into SLA doctoral students' statistical literacy and quantitative research methods training in SLA programs is timely and necessary. Indeed, as Jones (2013) highlighted, doctoral students are "the potential backbone of all research programs and, as such, are instrumental in the discovery and implementation of new knowledge" (p. 99). Given all these, in this study I investigate the statistical knowledge of the SLA doctoral students by using a statistics background questionnaire and a statistical literacy assessment survey designed to directly measure SLA researchers' ability to understand and interpret statistical analyses. Moreover, I use semi-structured interviews to further investigate doctoral students' experiences and training in quantitative analysis in light of the surveys.

1.1 The Use of Statistics in SLA

Although a variety of research methods are used by SLA researchers, several researchers have highlighted that quantitative research methods predominate L2 research and continue to increase in both complexity and sophistication (e.g., Gass, 2009;

Lazaraton, 2000, 2005; Norris et al., 2015; Plonsky, 2011, 2013, 2015). As is true of all fields that employ quantitative methods, statistics play a crucial role in analyzing data. Indeed, the use of statistics in SLA research has increased over the past 30-40 years (Brown, 2004; Loewen & Gass, 2009). In other words, most L2 research today relies on statistics in some form or another. For instance, in an attempt to provide a snapshot of the methodological culture of L2 research, Lazaraton (2000) reviewed 332 studies published in four different SLA journals (i.e., Language Learning, The Modern Language Journal, Studies in Second Language Acquisition and TESOL Quarterly). She found that 88% of these articles were quantitative in nature and the authors of them primarily used simple statistics such as *t* tests and ANOVAs. In a similar study, Lazaraton (2005) reviewed 524 articles in the same journals again and noted a similar amount of use of quantitative analysis (86%). This survey also indicated that between 2000 and 2005, most quantitative researchers began to employ a wider range of statistical procedures including descriptive statistics, ANOVAs, *t* tests, correlations, regression analyses, and chi-square tests. Most recently, Gass (2009) surveyed the types of data analyses, measures, and statistics that were used in L2 research and published across four different journals. She noted that the field has become “more sophisticated in its use of statistics” (p. 19). Indeed, despite their reliance on common parametric tests such as *t* tests and ANOVAs, researchers in the field have also begun to employ novel and more robust statistical techniques (Cunnings, 2012; Larson-Hall, 2010; Plonsky, Egbert & Laflair, 2014). For instance, methodological surveys have shown that some advanced statistical techniques such as confirmatory factor analysis, exploratory factor analysis and structural equation modeling have been applied considerably more frequently in L2 research, even if they are still not as common as

parametric tests (Plonsky 2011, 2015; Winke, 2014). In addition, several L2 researchers (e.g., Larson-Hall, 2010; Plonsky et al., 2014) have recommended that researchers use more robust statistics, such as bootstrapping, for small and non-normally distributed data sets, which are prevalent in L2 research. Another novel data analysis technique that has recently appeared in SLA research is mixed-effects modeling, which, in fact, has been employed in sub-domains of SLA such as language assessment and testing, and psycholinguistics (Cunnings, 2012; Linck & Cunnings, 2015). Using mixed-effects models enables L2 researchers to simultaneously investigate “participant-level and item-level factors in a single analysis” (p. 379), and can be of importance in longitudinal designs in L2 research.

Along with the current trend towards the use of novel and more sophisticated statistical methods in L2 research, there are an increasing number of discipline-specific statistics sources (e.g., books, articles and editorial comments) to which L2 researchers can refer. The first in-house instruction on statistical analyses using SPSS is Larson-Hall’s (2010, 2015) *A Guide to Doing Statistics in Second Language Research Using SPSS*. It provided a thorough explanation of basic descriptive and common inferential statistics. Another important example of the recent book-length methodological treatments is Plonsky’s (2015) edited volume, titled *Advancing Quantitative Methods in Second Language Research*, which covered some advanced yet under-used statistical concepts and procedures such as mixed-effects models, cluster analysis, discriminant function analysis, Rasch analysis, and Bayesian models. Such sources are definitely crucial in expanding the statistical repertoire of both consumers and producers of L2 research, and in keeping them up-to-date in their research areas.

Taken together, the results of the methodological surveys and the introduction of new publications devoted to quantitative research methods accentuate the significance and predominance of statistical procedures in the SLA field. As can be expected, the increased use of statistical procedures has led to the increased awareness of methodological issues, which I deal with in the following section.

1.2 Methodological Quality in SLA

It is important to adhere to rigorous research and reporting practices when conducting research because, as Gass, Fleck, Leder and Sveticks (1998) noted, “respect for the field of SLA can come only through sound scientific progress” (p. 407). As L2 researchers begin using more statistical techniques, journal editors and those who monitor research in the L2 field have an increased awareness of and concern for the quality of the techniques that are used. Indeed, a significant number of SLA researchers have drawn attention to the quality of statistical knowledge and methodology in the field (e.g., Godfroid & Spino, 2015; Larson-Hall & Plonsky, 2015; Loewen et al., 2014; Norris et al., 2015; Plonsky, 2011, 2013, 2015; Plonsky & Gonulal, 2015; Winke, 2014). These studies, mostly synthetic in nature, were written because the researchers sought to evaluate methodological quality in L2 research, and they primarily addressed the following issues: (a) study design, (b) instrumentation, (c) statistical analyses, and (d) reporting practices. Plonsky (2013) defined (methodological) quality as “the combination of (a) adherence to standards of contextually appropriate, methodological rigor in research practices and (b) transparent and complete reporting of such practices” (p. 657). These synthetic studies on methodological practices collectively point out that most quantitative L2 research falls short in at least one aspect of reaching high-methodological quality.

For instance, with respect to study design, one notorious problem in SLA research is sample size (Chaudron, 2001; Larson-Hall & Herrington, 2010; Plonsky & Gass, 2011). The sample size in L2 research tends to be small (generally less than 20, Plonsky, 2013), which creates a problem for statistical power. At the same time, there may be a tension between experimental rigor and the ecological validity of classroom-based research (Loewen & Plonsky, 2015). To illustrate, the total number of students in first year Turkish courses that I taught over five semesters is 16. Therefore, in such studies looking at less commonly taught languages, the criticisms related to low statistical power should be weighed against the ecological validity of the small samples, and using intact classes can be ecologically valid. Similarly, another problem that is very common, yet probably difficult to avoid, is the lack of true randomization in group selections (Larson-Hall; 2010; Larson-Hall & Herrington, 2009). In others words, samples in L2 research are mostly convenience-based. They are often based on treatments applied to intact classes, which is ecologically sound and practical, but not robust for scientific, empirical inquiry.

Related to statistical analyses, the most frequent problems found in L2 research include (a) overuse of some basic statistical tests (when more informative and robust statistics could have been applied instead), (b) frequently violated statistical assumptions, and (c) omission of non-statistical results (Chaudron, 2001; Norris, 2015; Norris et al., 2015; Plonsky, 2013; Plonsky & Gonulal, 2015). Related to the overuse of certain statistical tests, Brown (2015) noted that some researchers might be “stuck in a statistical rut” (p. 19), and thus tend to exclusively use a statistical method (probably the one they know very well) for a number of studies. Given that, he suggested that L2 researchers broaden their knowledge of statistical methods. Furthermore, Plonsky (2011, 2013) noted

that poor reporting practice is another common issue among L2 researchers. For instance, researchers tend to fail to report data crucial for readers to be able interpret the results and use them in subsequent analyses (see also Polio & Gass, 1997). As such, Plonsky and Gass (2011) and Larson-Hall and Plonsky (2015) suggested that L2 researchers at least report basic descriptive statistics, along with effect sizes and statistical power.

Methodological syntheses in SLA research are a recent yet up-and-coming research area where problems related to statistical analyses and data reporting can be detected and addressed. In recent years, a few researchers have urged caution and revision in current quantitative practices regarding certain statistical methods. To my knowledge, Plonsky and Gonulal (2015), and Winke (2014) are the first that took a meta-analytic approach to investigate the use of certain advanced statistical methods in L2 research. Plonsky and Gonulal (2015) investigated how exploratory factor analysis (EFA), a special type of factor analysis, is used by L2 researchers. Another purpose of the study was to discuss and illustrate how such types of methodological syntheses could contribute to the field. Plonsky and Gonulal reviewed and critically evaluated 51 EFA studies published in six different journals. The results showed that SLA researchers had several issues with following and/or reporting the necessary steps (e.g., factorability of the data, factor retention, extraction and rotation methods) of exploratory factor analytic procedures. In another methodological synthesis, Winke (2014) investigated the extent to which SLA researchers adhered to standards of methodological rigor when carrying out structural equation modeling (SEM). Winke examined 39 SEM studies published between 2008 and 2013. The results indicated that although SEM was well applied in

several studies, four areas (i.e., sample size, model presentation, reliability, and Likert-scale points) appeared to be recurrently problematic.

In addition, the status quo of statistics in SLA has recently received much editorial and scholarly attention. For instance, the authors of the 2015 volume of the *Currents in Language Learning* series have specifically focused on enhancing the statistical literacy, thinking and reasoning in the field of SLA by addressing common issues, challenges and proposed solutions along with important advances in quantitative research. Indeed, such a volume is important and timely in pinpointing the state of current quantitative research practice in the field and the need for improvements in methodological training in graduate programs, to which I return below.

1.3 Graduate Training in Quantitative Research

When viewed in its entirety, the use and variety of statistics in L2 research seems less than optimal, even though it has increased over the years. A number of researchers (Aiken et al., 2008; Capraro & Thompson, 2008; Henson et al., 2010) in neighboring disciplines such as education and psychology argued that the ability of researchers to conduct high-quality research is influenced by the quality of the methodological training they receive. For instance, Henson et al. (2010) asserted that there is a close relationship between statistical training and the application of quantitative research methods in published scholarly work, although there might be some other possible factors. Yet, as Thompson (1999) highlighted, doctoral curricula in many disciplines “seemingly have less and less room for quantitative, statistics, and measurement content, even while our knowledge base in these areas is burgeoning” (p. 24). Further, given the fact that research methodology is a dynamic field that sees regular improvements in statistical procedures

(Norris, 2015; Skidmore & Thompson, 2010), investigation of methodological training in SLA doctoral programs appears to be necessary. This area of research has drawn more scholarly attention in other fields than in SLA. There have been several studies in which the authors explored research methodology curriculum in fields such as psychology (Aiken et al., 1990, 2008; Zimiles, 2009), education (Leech & Goodwin, 2008), and sub-fields such as counselor education (Borders et al., 2014) and educational statistics (Curtis & Harwell, 1998). For instance, Leech and Goodwin (2008) investigated the research methods course requirements in 100 education doctoral programs. The mean number of required methods courses was 3.67 ($SD = 1.91$). Leech and Goodwin found that most programs (62%) required students to take a quantitative research methods course. At a closer look, 63% of education programs required basic statistics and 54% intermediate statistics. In a recent study, Borders et al. (2014) reviewed research training in 38 counseling doctoral programs. They found that although the range of statistical training offerings varied, most counseling programs provided a thorough coverage of basic descriptive statistics and common inferential statistics but not new and more sophisticated statistics. In addition to this finding, the researchers noted that because most of the quantitative research methods courses were offered outside the counseling programs, these courses tended to lack relevance for typical research conducted in counseling. Border et al. reported that only half of the faculty (58%) had positive feelings about their research training whereas almost 20% were not pleased with their research training. It should be noted, however, that these studies on doctoral training in statistics did not include an exploration of the competence of students in statistics, but counted on faculty impressions of the adequacy of statistical training.

Ostensibly, research on methodological training has gained momentum in other fields. It is surprising that little has been written regarding what parts of graduate programs in SLA are devoted to quantitative research methods. What SLA researchers and applied linguists know about the current content and nature of graduate training in quantitative research methods in the field is largely limited to a few studies (e.g., Bailey & Brown, 1996; Brown, 2013; Brown & Bailey, 2008; Gonulal et al., in preparation; Lazaraton et al., 1987; Loewen et al., 2014). Brown and Bailey (2008), a recent replication of Bailey and Brown (1996), investigated the language testing course instructors' backgrounds, the content and structure of language testing courses, along with students' attitudes towards such courses. The results showed that most language testing courses covered common item statistics (e.g., *item facility*, *item discrimination*, *item quality analysis*), test reliability estimate methods (e.g., *test-retest reliability*, *parallel forms reliability*, *inter-rater* and *intra-rater reliability*), test validity methods (e.g., *content validity*, *construct validity*, and *criterion-related validity*) and descriptive statistics, whereas some other more sophisticated statistics (e.g., *biserial correlation*, *Rasch analysis*, *split-half method*, *K-R20*, *K-R21*, *Spearman-Brown prophecy formula*, *Kappa generalizability coefficient*) were not covered in 25% to 68% of the courses. As for students' attitudes towards language testing courses, approximately 70% found the courses interesting and useful while roughly 35% found the courses difficult and 13% highly theoretical.

When looking at the field from a broader perspective, Loewen et al. (2014) reported that the average number of quantitative research methods courses taken by SLA graduate students is two, with most courses in education departments, followed by

applied linguistics and SLA departments. In a more recent study, Gonulal et al. (in preparation) investigated the development of statistical knowledge among SLA graduate students. In particular, the researchers attempted to explore the potential gains in statistical knowledge made by a group of SLA graduate students including both master's and doctoral students at four American universities during semester-long discipline specific statistics courses (i.e., introduction to quantitative research methods and intermediate statistics). The results showed that students increased their knowledge of basic descriptive statistics and particularly, common inferential statistics, with the highest gains being reported for *degrees of freedom, statistical power, post hoc tests, ANOVA* and *effect size* whereas the lowest gains were on *Rasch analysis, SEM, and factor analysis*. Understandably, the students' knowledge base concerning common inferential statistics had more room for growth because students had already some basic statistical knowledge at the beginning of the course. These results also indicated that although the existing statistical training in the field may not reflect some of the advances in statistical analyses (e.g., *factor analysis, bootstrapping, SEM, mixed-effects models*), it is still gratifying to see that some of the recent critiques in statistical analyses (e.g., *statistical power, effect size*; see Gass & Plonsky, 2011, Larson-Hall & Plonsky, 2015) are finding their way into the content of statistical training in the field.

Besides the content and amount of statistics courses offered in the field of SLA, it is equally important to focus on the strategies to teach statistics. Unfortunately, the literature on teaching statistics in SLA programs is mostly limited to Brown's (2013) commentary on language testing courses. In looking at the general literature, most studies on teaching statistics are not empirical but "largely anecdotal and comprises mainly

recommendations for instruction based on the experiences and intuitions of individual instructors” (p. 71, Becker, 1996). Indeed, a variety of strategies (as cited in Brown, 2013) have been proposed to effectively teach statistics: (a) *need-to-know approach* (Fischer, 1996) deals with what students should be able to do with statistics, (b) *reasoning-from-data approach* (Ridgeway, Nicholson, & McCusker, 2007) draws on mostly on statistical reasoning, (c) *real data approach* (Singer & Willet, 1990) and (d) *linking statistics to the real world approach* (Yilmaz, 1996), both of which include using real data sets so that students students can apply what they learn to their own research. Although these strategies look promising, they need to be further investigated. Overall, as can be seen, a complete picture of what research methods courses are being offered in SLA programs, what is taught what kinds of teaching strategies are used in these courses is still lacking.

Of course, it is important to note here that one can improve his or her statistical knowledge through different routes. *Self-instruction* and self-training are two, closely similar yet different, ways. When looking at the definition of self-instruction, different researchers have defined it in different ways in different contexts. In one of the earlier definitions, self-instruction was defined as “situations in which a learner, with others, or alone, is working without the direct control of a teacher” (Dickinson, 1987, p. 5). Similarly, Jones (1998) defined it “a deliberate long-term learning project instigated, planned, and carried out by the learner alone, without teacher intervention” (p. 378). Even though there is no clear definition of self-training, what it means and encompasses seems to be somewhat broader. For instance, although a workshop may not count as self-instruction, it may count as self-training. That is, self-training not only contain self-

teaching but also self-regulated learning which may include expert-led learning in a non-required pedagogical environment. Although, to my knowledge, no studies have investigated the effects of self-training in learning statistics, Rossen and Oakland (2008) anecdotally noted that it is possible for students to maintain and improve their knowledge of statistics through external, additional and self-paced statistical training. However, Golinski and Cribbie (2009) argued against this claim, anecdotally stating “in our opinion, it is unlikely that a significant number of psychology students are gaining extensive knowledge in quantitative methods in a self-taught manner” (p. 84). Considering these opposing views, further research and clarification are needed in this area.

1.4 Statistical Literacy

As the field is becoming “more sophisticated in its use of statistics” (Gass, 2009, p. 19), several methodological issues (e.g., inappropriate use and overuse of certain statistical methods or poor reporting practices) have arisen. Several researchers (e.g., Norris et al., 2015; Plonsky, 2013) attributed some of these methodological quality problems to the limited state of statistical literacy among L2 researchers. Given the predominance of quantitative studies in L2 research, statistical literacy appears to be a critical skill to acquire on the parts of both the producers and consumers of L2 research. Statistical literacy is a new research area in L2 research, although it has been investigated in other fields, mostly in statistics and mathematics education. In the following two sections, I provide definitions of statistical literacy and other different, yet, related terms, and then look at the studies conducted to measure statistical literacy.

1.4.1 Statistical literacy and other related terms

Before grappling with the definitions of statistical literacy, it is necessary to first start with the concept of *literacy*. The American heritage dictionary of the English language defines literacy as “the ability to read and write, and the condition or quality of being knowledgeable in a particular subject or field” (online version). Dauzat and Dauzat (1977) also provided a similar definition where literacy is again described as “the ability to read and write in a language”, emphasizing that it is not “an all or none proposition” but includes various levels (p. 40). As for a broader view of literacy, the national literacy act defined literacy as “an individual’s ability to read, write and speak in English, and compute and solve problems at a level of proficiency necessary to function on the job and in society, to achieve one’s goals, and develop one’s knowledge and potential” (as cited in Kirsch et al., 1993, p. 28). Over the years, the concept of literacy has expanded to various areas, and now there are various types of literacy including computer literacy, cultural literacy, digital literacy, information and statistical literacy.

Statistical literacy, with different terms and expressions (e.g., statistical reasoning, statistical thinking), has been focused on in different fields as the fields push to improve the ability of people to consume and produce data. Just as in definitions of literacy in general, different definitions of statistical literacy have been proposed. One of the earlier descriptions of statistical literacy was provided by Wallman (1993):

“Statistical Literacy” is the ability to understand and critically evaluate statistical results that permeate our daily lives—coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions (p. 1).

In line with the definition of Wallman, Watson (1997) introduced a three-layered definition of statistical literacy with increasing sophistication: (a) ability to understand basic statistical concepts, (b) ability to understand statistical terminology and concepts embedded in a broader social context, (c) ability to challenge or critically evaluate statistical information in media. In the same way, Schield (1999, 2004) emphasized that statistical literacy means more than number crunching in that statistically literate individuals should be able to understand what is being asserted, think critically about statistical arguments, and have an inductive reasoning about such arguments.

In another comprehensive study on statistical literacy, Gal (2002) defined statistical literacy focusing on two broad but related parts:

(a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions (pp. 2-3).

Further, Gal also proposed a model of statistical literacy that centers mostly on consumers of data. His model comprises two primary components: a) a knowledge component, which includes literacy skills, mathematical knowledge, statistical knowledge, context knowledge and critical questions, and b) a dispositional component including beliefs and attitudes, and critical stance. When looking closely at the elements in each component, since most statistical information is presented through written or oral texts or in graphical format, Gal considered literacy skills as prerequisite for statistical literacy because limited literacy skills may impede skills important for statistical literacy. In addition, according to Gal, individuals should have some basic understanding of

mathematical procedures used in some common statistical concepts such as percent, mean and median.

As for the statistical knowledge element of statistical literacy, Gal (2002) divided statistical knowledge into five sub-components: “(a) knowing why data are needed and how data can be produced, (b) familiarity with basic terms and ideas related to descriptive statistics, (c) familiarity with graphical and tabular data and their interpretation, (d) understanding of basic notions of probability, and (e) knowing how statistical conclusions or inferences are reached” (p. 10). According to Gal, apart from mathematical and statistical knowledge, context knowledge is also important because appropriate interpretation of statistical information can be affected by an individual’s familiarity with the context where the statistical information is embedded. The final knowledge element of statistical literacy pertains the ability to critically evaluate statistical messages. As much similar to critical questions element, which is another aspect of knowledge component, the dispositional component of Gal’s statistical literacy refers to the propensity to have a questioning attitude towards statistical messages.

Considering all these definitions, it appears that statistical literacy entails a sophisticated way of looking at statistical information. Another common theme among these definitions is that statistical literacy focuses mostly on data consumers. In fact, in a more recent definition, Schield (2010) distinguished statistical literacy from statistical competence in that the former addresses data consumers whereas the latter is a necessary ability for data producers.

Statistical reasoning and statistical thinking are two other frequently used terms related to statistical literacy. Although statistical literacy and statistical reasoning are

often used interchangeably, several researchers (e.g., Ben-Zvi & Garfield, 2004; Garfield, 2003; Garfield & Ben-Zvi, 2007) considered statistical reasoning as a step after statistical literacy, with statistical literacy considered a basic but important ability to understand basic statistical concepts and terminologies. According to Garfield and her colleagues, statistical reasoning includes both the ability to understand and explain statistical procedures, and the ability to fully interpret statistical messages. However, statistical thinking is a marginally more inclusive term embracing not only statistical literacy but also statistical reasoning (Wild & Pfannkuch, 1999). In line with Wild and Pfannkuch's (1999) explanation, Ben-Zvi and Garfield (2004), and Garfield and Ben-Zvi (2007) argued that when compared to the other two concepts, statistical thinking requires a slightly more sophisticated way of thinking. In more concrete terms, statistical thinking is similar to having a mindset of a statistician in that it refers to "the knowing how and why to use a particular method, measure, design or statistical model; deep understanding of the theories underlying statistical processes and methods as well as understanding the constraints and limitations of statistics and statistical inference" (Garfield & Ben-Zvi, 2007, p. 381).

In considering all these, there is no unanimity in the definitions of statistical literacy, statistical reasoning and statistical thinking, probably because they are highly interrelated. Following key points from all these definitions, I operationalized statistical literacy within the domain of SLA as the ability to (a) understand basic statistical terminology, (b) use statistical methods appropriately, and (c) interpret statistical analyses, which may be encountered in L2 research contexts (I will revisit this definition later in the discussion chapter).

In the following sections, I focus on how to assess statistical literacy, in light of previous statistical literacy assessment studies conducted mostly in statistics and mathematics education.

1.4.2 Research on statistical literacy

Assessment of statistical literacy can be done in several ways, such as written and oral exams, formative and summative assessments, and large-scale assessments. When looking at the design and type of tasks in statistical literacy assessment, Watson (1997) considered context as a vital element. In addition, Schield (2010) provided four ways to assess statistical literacy. These ways included asking students to (a) evaluate the use of statistics in a real-life data set, (b) calculate a quantity or make a statistical judgment in a given scenario, (c) understand and interpret statistical information presented in a graphical or tabular format, and (d) answer multiple-choice questions on certain statistical concepts and procedures. With the increased interest in statistical literacy, several statistical literacy instruments (e.g., *Statistical Literacy Inventory*, *Statistical Reasoning Assessment* and *Statistics Concepts Inventory*) have been developed, measuring statistical literacy in at least one of these ways.

Schild's (2002) *Statistical Literacy Inventory* (SLI) is one of the statistical literacy assessment surveys designed to measure statistical literacy. The SLI includes 69 items focusing on reading and interpreting percentages and rates presented in tabular and graphical format. Of 69 items, 63 include three response options (i.e., yes, no and don't know) and the last 6 items related to evaluation on the SLI includes four options (i.e., from strongly agree to strongly disagree). However, this survey appears to be more appropriate for assessing the statistical literacy of citizens.

In their study investigating the construct of statistical literacy, Watson and Callingham (2003) used an 80-item statistical literacy instrument designed for students in grades 3 through 9. The instrument included open-ended questions focusing on sampling, average, variation, chance, and graphs, using a 4-point coding system.

Another instrument is the Statistical Reasoning Assessment (SRA) designed by Garfield (2003). As its name suggests, the SRA focused on assessing statistical reasoning. The SRA consisted of 20 multiple-choice items, most of which also included sub-questions asking participants to provide a rationale for their choice. The SRA was used with students in high school and college level statistics courses to investigate their reasoning about sample, population, types of variable (e.g., discrete, continuous), measures of center and spread, correlation and probability.

The other related instrument is the Statistics Concepts Inventory (SCI) assessing engineering and mathematics students' conceptual understanding of fundamental statistics. This multiple-choice survey had initially 38 items but Allen (2006) modified the survey and proposed a shorter version, with 25 items. The SCI included four sections (i.e., descriptive, probability, inferential and graphical) covering a variety of statistical concepts and procedures such as descriptive statistics, probability distributions, correlations, parameter estimation, linear regression, type I and type II errors, and Bayes' theorem.

In considering these statistical literacy instruments, a couple points stand out. First, the instruments were designed for different age groups. Moreover, the content and scope of the instruments changes from a few simple statistics to a number of inferential statistics. Further, some instruments (e.g., SCI) appear to be field-specific. That is, the

items on the instruments were contextualized for certain fields. Therefore, these instruments are not directly applicable to the field of SLA (I will return to this point later in the method chapter).

When looking at the studies conducted to assess statistical literacy of college students or adults, Schield (2006) conducted a study using the SLI instrument with 169 adults including U.S. college students ($N = 85$), college teachers worldwide ($N = 43$) and data analysts in the United States and South Africa ($N = 47$). In his study, Schield focused on measuring participants' ability to understand and interpret simple statistics (i.e., percentages and rates) presented in tabular and graphical formats. A great number of college teachers (78%) and data analysts (87%) had taken at least one statistics course while approximately one third of them (29% of college teachers and 34% of data analysts) had taken at least two courses. However, the number of statistics courses college students had taken was not reported. The results showed that the mean error rate for college students was 50%, for data analysts was 45% and for college teachers was 30%. The highest error rates were on items related to interpretation of tabular and graphical data. Schield called for a need to teach statistical literacy or rather enhance statistical literacy through statistics courses at college levels.

In a relatively different context, Galesic and Garcia-Retamero (2010) conducted a cross-cultural study between Germany and the United States to investigate statistical knowledge (i.e., probability and chance) of approximately 2000 adults (1001 adults from Germany, and 1009 from the United States) within a medical context. The researchers used a 9-item, short answer statistical numeracy scale. The results indicated participants from Germany and the United States performed similarly on the statistical numeracy

scale. That is, German participants answered 68.5% of the items correctly and American participants answered 64.5% of the items correctly. Galesic and Garcia concluded that physicians should not take for granted that patients can easily comprehend basic medical-related statistics (e.g., probability and chance) used to express the advantages and disadvantages of medical treatments.

In a more recent study, Pierce and Chick (2013) conducted a mixed methods study to investigate 704 Australian school teachers' attitudes towards box-plot data and their ability to interpret such graphical data. The results showed that although teachers had positive feelings towards using graphical data representation methods (i.e., box-plots) unlike tabular data representations, some reported that they found such graphical data hard to interpret. Indeed, Pierce and Chick found that most school teachers could interpret box-plots "at a superficial level" (p. 203).

Overall, although these studies on statistical literacy worked with different participant profiles, it seems that participants had some issues when interpreting statistical information. In line with this interesting point, Norris (2015) argued that the use and interpretation of significance tests in the field of SLA is problematic. In the following section, I focus on SLA-specific statistical literacy studies, though there are not many (3 in total, to my knowledge).

1.4.3 Statistical literacy in SLA

In spite of the apparent significance of statistical literacy as a necessary skill to be acquired by SLA researchers, very few studies on the state of statistical literacy among L2 researchers exist. The first comprehensive study investigating L2 researchers' statistical literacy and attitudes was conducted by Lazaraton et al. (1987). They had 121

professionals in applied linguistics completed a comprehensive statistical literacy survey. Participants self-rated their degree of familiarity with 23 statistical concepts and procedures, and to respond to 18 statements regarding attitudes towards statistics and quantitative research methods. The results indicated that the participants were comfortable in interpreting and using some basic statistical concepts and procedures such as *mean*, *median*, *null hypothesis*, *validity*, *reliability*, *standard deviation*, whereas they were less confident with some of the comparatively more advanced statistical methods and procedures such as *implicational scaling*, *power analysis*, and *Scheffé test*. Although there were varying attitudes, participants mostly agreed that statistical literacy is a necessary skill and thus L2 researchers should take a research design/statistics course. In a similar but more recent study, Loewen et al. (2014) conducted a study looking at the statistical knowledge of 331 applied linguists and SLA researchers, including both graduate students and professors, in a partial replication of Lazaraton et al.'s study. The result echoed the findings of Lazaraton et al. in that statistical literacy was found to be a necessary component of L2 research. Further, L2 researchers' attitudes towards statistics and quantitative research were largely positive. Loewen et al. also investigated the predictors of statistical self-efficacy and attitudes towards statistics. They found that number of statistics courses an individual took and quantitative research orientation were predictive of attitudes towards statistics and statistical self-efficacy.

Although these two studies are valuable in providing a snapshot of the statistical literacy in the field, the researchers of both studies relied on self-report data and included two different groups of participants (i.e., faculty and graduate students). Therefore, to have more reliable information regarding the current state of statistical literacy among

graduate students, research that directly measures researchers' ability to use and interpret statistical methods is much needed, following other similar studies (e.g., Schield, 2002; Pierce & Chick, 2013) that were conducted to measure statistical literacy in other fields. A discipline-specific instrument measuring researchers' knowledge in statistics should be developed because researchers' ability to interpret and use statistical procedures might be different from what researchers assume they can do: They might over- or underestimate their ability. Such an instrument should be able to assess researchers' actual knowledge, reasoning, thinking, and conceptual understanding of statistics within the context of SLA.

1.5 Research Questions

Given the strong quantitative research tradition in the field of SLA, being statistically literate is important, not only for producers of but also for consumers of SLA research. In order to accurately inform L2 theory and practice, SLA researchers, particularly newly minted researchers, need to ensure that they are conducting and reporting statistical analyses properly. However, as can be seen in several methodological studies (e.g., Larson-Hall, 2010; Larson-Hall & Herrington, 2010; Norris, 2015; Plonsky, 2011, 2013; Plonsky & Gonulal, 2015; Winke, 2014), most L2 researchers that apply statistics, sophisticated and novel statistics in particular, fall short in at least one aspect of reaching high-methodological quality. Indeed, the current state of methodological quality of L2 research is closely related to the level of statistical literacy among L2 researchers. Although a few studies (i.e., Gonulal, et al., in preparation; Lazaraton et al., 1987; Loewen et al., 2014) have been conducted to capture the current state of statistical literacy in L2 research, there remains a paucity of evidence on how statistically literate SLA doctoral students are in the field. The importance of statistical literacy, taken

together with the dearth of evidence of SLA doctoral students' ability to understand and interpret quantitative L2 research, was the impetus of this study. This study is novel in several ways. This research project is an initial attempt to develop a discipline-specific instrument targeting SLA researchers' statistical literacy. With the present study, I aim to provide some direct evidence of SLA doctoral students' ability to understand and interpret statistical analyses. In addition, this study will shed light on the status of statistical training among doctoral students in SLA in North America. The following research questions guided my study:

1. To what extent have SLA doctoral students received training in statistics?
2. How statistically literate are SLA doctoral students?
3. What kinds of variables predict SLA doctoral students' statistical literacy?
4. What are the general experiences and overall satisfaction of the statistical training of SLA doctoral students?

CHAPTER 2: METHOD

The purpose of this exploratory study was to provide a snapshot of SLA doctoral students' current state of statistical literacy, their statistical training and experiences with statistical analyses as well. In doing so, I used a concurrent or convergent mixed-methods research design (Creswell & Clark, 2011), which enabled me to collect different yet complementary data to adequately address the complex nature of statistical literacy. I used a variety of data collection methods such as surveys for quantitative data, and semi-structured interviews, comments left at the end of the survey and some e-mail exchanges for qualitative data. In this chapter, I provide detailed information about the participants who participated in the study and the instruments that I used. Then, I give details regarding the statistical analyses I performed.

2.1 Participants

Participants were graduate students pursuing a doctoral degree in SLA, second language studies, applied linguistics or related programs in North America. Due to the potential differences in graduate training between the programs in North America and the rest of the world, I limited the scope of the study to North America. Of the approximately 900 graduate students that I was able to reach out, 125 took the *SLA for SLA* survey (I will explain the survey in detail later in this chapter). However, 5 participants were excluded from the analyses since they reported to have used additional sources (e.g., statistical textbooks, internet) when answering the survey questions, which left the sample size at 120. Of these 16 participated in follow-up semi-structured interviews. The

participants were from thirty universities across North America (see Appendix A for the list of the universities from which the participants were recruited).

Figure 1 below shows the geographic location of the participants. It is a color-coded map of the United States of America and Canada based on the number of participants who participated in the study from the different locations ($N = 108$; 12 participants did not mark the location of their institution on the map). The color changes from dark blue to red depending on the number of participants in a certain state (dark blue represents 1 participant; red represents 11 participants). Overall, given the fact that this study included participants from a wide range of locations in North America, the current sample appeared to be representative of the target population of the present study: North American doctoral students in SLA.

There were 74 females and 46 males, whose ages ranged from 24 to 42 ($M = 30.82$, $SD = 3.95$). Participants were in different years of their doctoral program. 18% were first-year, 25% second-year, 26% third-year, 15% fourth-year graduate students. 16% of the participants were in their fifth year or more. Approximately half of the participants (47%) were in an SLA program, followed by applied linguistics (27%), TESOL/TEFL (12%), language testing (4%), foreign languages (3%), and other programs (8%) such as psycholinguistics, corpus linguistics, and English.

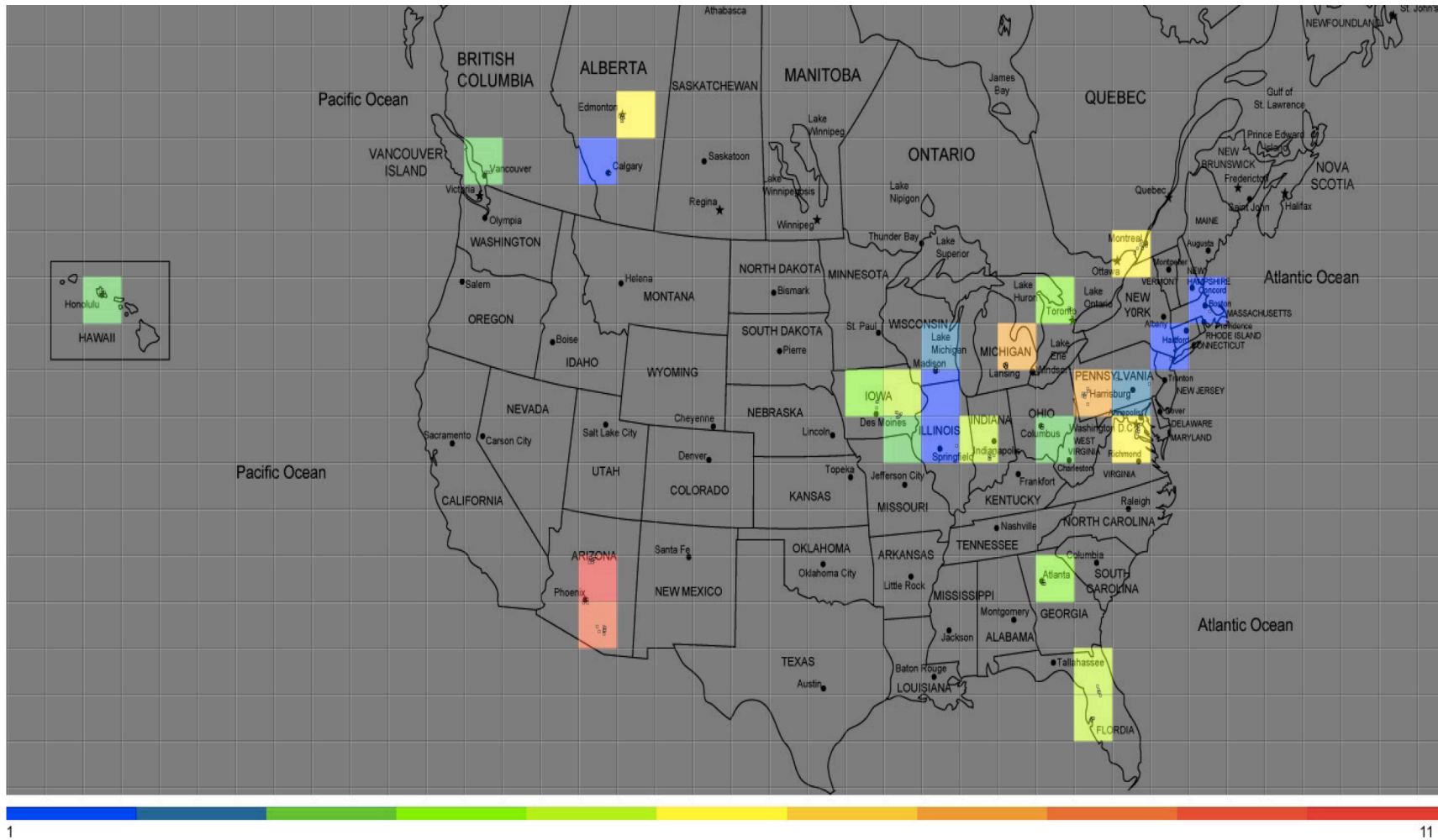


Figure 1. Geographic information about the participants

2.2 Instruments

Data for this study came from three major sources: (a) a statistical background questionnaire, (b) a statistical literacy assessment survey, and (c) semi-structured interviews. Apart from these sources, I also had e-mail exchanges with two graduate students who neither took the survey nor participated in the follow-up interviews but shared their opinions about the study.

2.2.1 Statistical background questionnaire

In order to elicit information about participants' statistical training, I developed this questionnaire closely based on Loewen et al.'s (2014) questionnaire. Along with basic demographic questions, the questionnaire consisted of 10 items addressing participants' research orientation, the number of statistics courses taken, the departments that those statistics courses were taken, the amount of statistical training, the amount of self-training in statistics, the types of statistical assistance participants tended to seek, the software programs used to calculate statistics, and self-rated statistical literacy (see Appendix B).

2.2.2. Development of a discipline-specific statistical literacy assessment

Given that there is no unanimous definition of statistical literacy in the literature, it was not surprising to see that there was no all-encompassing assessment instrument of statistical literacy. There were several statistical literacy assessment instruments (e.g., Statistics Concept Inventory [SCI], Statistical Literacy Inventory [SLI] and Statistical Reasoning Assessment [SRA]) specifically designed to assess either the learning

outcomes in introductory-level statistics courses or the general use of informal statistics in everyday life.

You are rolling dice. You roll 2 dice and compute the mean of the numbers rolled, then 6 dice and compute the mean, then 10 dice and compute the mean. Under which scenario would you be most surprised to find a mean of at least 4.5?

- a) Rolling 2 dice
- b) Rolling 6 dice
- c) Rolling 10 dice
- d) There is no way this can happen

Correct Answer: C
Topic Area: Probability

Figure 2. Example item on the Statistics Concept Inventory (Allen, 2006, p. 433).

SMOKERS

Religion	Percentage
Catholics	20%
Protestants	40%
Other	40%

Do you think the following statements accurately describe the data shown in this pie chart?

9. **20% of smokers are Catholics.**
a. Yes b. No c. Don't know

10. **Protestants (40%) are twice as likely to be smokers as are Catholics (20%).**
a. Yes b. No c. Don't know

Figure 3. Example item on the Statistical Literacy Inventory (Schild, 2002, p. 2).

However, not surprisingly, these instruments are not completely applicable to researchers in the field of SLA because those instruments had items (e.g., mathematical

calculations, permutations, combinations, conditional probabilities, see a sample item in Figure 2) that were not necessarily relevant to SLA researchers and research, or were more appropriate for certain groups such as mathematics and engineering students (e.g., SCI instrument, Allen, 2006) or a broader group (e.g., Schield's SLI for citizens, see a sample item in Figure 3). As Gal (2002) and Watson (1997) highlighted, context in statistical literacy assessment is critical because the context in which statistical information is presented is the source of meaning and basis for interpretation of statistical results.

2.2.2.1 Statistical literacy assessment for second language acquisition survey

Given that there was no established instrument that can measure statistical literacy in the field of SLA, it was time to create a discipline-specific statistical literacy assessment instrument to investigate the statistics knowledge of SLA researchers. The statistical literacy assessment for second language acquisition (SLA for SLA) instrument was originally created for an independent group research project (unpublished research project) investigating the statistical knowledge of SLA faculty. I and several other SLA doctoral students who are also the members of the *Donuts and Distribution Statistics Discussion Group* in the Second Language Studies program at Michigan State University designed the *SLA for SLA* instrument under the supervision of Dr. Shawn Loewen. Drawing mostly on the definitions of Watson (1997, 2011) and Gal (2002), we came up with a working definition of statistical literacy for the project. We defined statistical literacy within the domain of SLA as the ability to understand, use and interpret statistical information typically encountered in L2 research. Following our definition of statistical literacy, we designed the survey to measure the ability to (a) understand basic statistical

terminology, (b) use statistical methods appropriately, and (c) interpret statistical analyses properly.

Table 1

Current statistics self-efficacy by Finney and Schraw (2003, p.183)

-
1. Identify the scale of measurement for a variable
 2. Interpret the probability value (p-value) from a statistical procedure
 3. Identify if a distribution is skewed when given the values of three measures of central tendency
 4. Select the correct statistical procedure to be used to answer a research question
 5. Interpret the results of a statistical procedure in terms of the research question
 6. Identify the factors that influence power
 7. Explain what the value of the standard deviation means in terms of the variable being measured
 8. Distinguish between a Type I error and a Type II error in hypothesis testing
 9. Explain what the numeric value of the standard error is measuring
 10. Distinguish between the objectives of descriptive versus inferential statistical procedures
 11. Distinguish between the information given by the three measures of central tendency
 12. Distinguish between a population parameter and a sample statistic
 13. Identify when the mean, median, and mode should be used as a measure of central tendency
 14. Explain the difference between a sampling distribution and a population distribution
-

The development of the *SLA for SLA* survey consisted of several phases. In the first phase, we designed the survey blueprint to outline the set of statistics concepts, procedures and tests that would be covered in the survey. To this end, we made use of a reliable and highly-cited statistics survey designed by Finney and Schraw (2003) as a guide during the development of the preliminary survey blueprint. This survey consisted of 14 items that ask about “confidence in one’s abilities to solve specific tasks related to statistics” (p. 164). As can be seen in Table 1, the items vary from distinguishing between

population and sample to interpreting the results of a statistical procedure. We used these items as the basis of the *SLA for SLA* blueprint.

In addition, since the content included in the *SLA for SLA* survey should be relevant to SLA researchers, we carefully reviewed several statistics syllabi collected from a variety of SLA and applied linguistics programs (e.g., Georgia State University, Georgetown University, Northern Arizona University, Michigan State University, and University of South Florida), and L2-oriented statistics textbooks (e.g., Larson-Hall, 2010; Mackey & Gass, 2015) to see to what extent the content domains addressed in Finney and Schraw's (2003) survey were covered in the field of SLA. For example, the topics that appeared to be less important (e.g., the difference between parameter and statistic, and probability rules) were not included. Instead, we included new items such as effect size. Further, we did not include advanced statistical topics (e.g., discriminant function analysis, mixed-effects regression models, structural equation modeling and Rasch analysis) on the survey because most SLA programs do not require their students to take advanced statistics courses that cover such topics. The second but probably more important reason was that we wanted to have a slightly shorter survey to reach doctoral students with different degrees of statistical inclination.

To identify question format and types used in such literacy studies, we also examined several statistical literacy instruments used in other fields (e.g., SCI, SRA) during the item development process. Taking all these important points into consideration, we initially created 35 multiple-choice items. Thirty of these items were based on nine L2-research related scenarios and 5 items were scenario-independent. In the next phases, the instrument went through several edits and changes. First, in order to

make the instrument more manageable, we decreased the number of scenarios from nine to five. This second version consisted of 30 multiple-choice items. Several SLA researchers reviewed several iterations of the second version for clarity.

Table 2

List of the content domains addressed in the SLA for SLA instrument

Skills	Items
1. Identifying the scale of measurement for a variable	Item 11, Item 17
2. Understanding of the difference between a sample and population	Item 1, Item 2
3. Understanding of the difference between descriptive and inferential statistics	Item 3, Item 20, Item 21
3. Distinguishing between the information given by the three measures of central	Item 22, Item 21
4. Explaining what the value of the standard deviation means in terms of the variable being measured	Item 4, Item 5,
5. Identifying if a distribution is skewed when given the values of three measures of central	Item 6
6. Ability to interpret a boxplot	Item 7, Item 8
7. Ability to select the correct statistical procedure to be used to answer a research question	Item 9, Item 10
8. Ability to interpret the results of a statistical procedure in terms of the research question	Item 12, Item 18, Item 26
9. Understanding of the difference between a Type I error and a Type II error	Item 13, Item 19, Item 27, Item 28
10. Understanding of power and effect size	Item 14, Item 24
11. Understanding of what the standard error means	Item 15, Item 16 Item 25

Then, the survey was reviewed by two SLA faculty with considerable quantitative research experience. We used the faculty members' detailed feedback to modify the instrument. The third version of the instrument consisted of five scenarios and twenty-eight multiple-choice questions related to these scenarios (see Table 2 for the structure of the *SLA for SLA* instrument). In addition, the instrument included sub-questions asking participants to give each item a rating ranging from 1 (not confident at all) to 10 (very

confident) to indicate participants' level of confidence in their response. Further, considering the possible attrition rate in a highly quantitatively-oriented study, we decided to randomize the scenarios in order to have a roughly similar number of responses for each item on the survey. However, we did not randomize the items within scenarios.

2.2.2.2 Pilot test

As a next step, we, the donuts and distribution statistics discussion group, piloted the third version of the instrument with 48 SLA faculty across North America. There were 28 females and 20 males, with a mean age of 42 years ($SD = 9.3$). Participants had different academic positions. Sixteen of them were assistant professors, 11 associate professors, 7 professors, and 13 had other positions (e.g., lecturer, language center director, writing instructor). Participants reported that they conducted quantitative research ($M = 3.95$, $SD = 1.53$) more frequently than qualitative research ($M = 3.04$, $SD = 1.46$), on a scale from 1 (*not at all*) to 6 (*exclusively*). In other words, the participants were slightly biased toward quantitatively-oriented research. Although the sample of the pilot test was different from the target population in this study, we originally designed the *SLA for SLA* survey to measure the statistical knowledge of SLA researchers including both faculty and graduate students. Therefore, we considered any information obtained from the pilot test regarding the survey valuable.

After collecting the data for the pilot test, I conducted an in-depth item analysis to examine the quality of the items on the survey. The overall reliability of the survey (Cronbach's $\alpha = .92$) was quite high. Furthermore, I performed both item-level and test-level analyses. More specifically, I calculated item difficulties, item discrimination

indices and confidence levels on the pilot data. An item difficulty index shows the percentage of participants who correctly answered the item. As its name suggests, item difficulty shows how difficult an item is. An item with an item difficulty value below .30 is usually considered very difficult whereas an item with an item difficulty value above .70 is considered easy (Brown, 2005). Item discrimination shows “how well an item discriminates between test takers who performed well from those who performed poorly on the test as a whole” (Brown, 2005, p.68). Items with low discrimination values (i.e., below .3) indicate that the items are not measuring the same construct as other items on the test or have wording issues, and thus may need to be revised or even dropped. Further, I examined participants’ confidence level scores as an another way of detecting the problematic items. For instance, when an item has a low item difficulty value (i.e., a difficult item) and a high mean confidence level, it can be interpreted that the answer key might be misleading for that particular item.

As can be seen in Figure 4 which presents some graphical information about the items on the survey based on their difficulty levels, item discrimination indices and confidence levels, several items (i.e., items 5, 9, 10, 14, 17 and 20c) had low or below cut-off level item difficulties and item discrimination, which indicated some problems on these items. In addition, confidence level scores indicated another potentially problematic item (i.e., item 16).

Along with these quantitative data, I also took a closer look at participants’ comments that they left at the end of the *SLA for SLA* survey regarding the design or any other aspects of the survey. Several participants commented on wording issues in a couple of items and made some useful suggestions such as explicitly stating the alpha

level for the scenarios. All in all, I used the information drawn from the analysis of the pilot data to modify the items on the final version of the *SLA for SLA* instrument (see Appendix C).

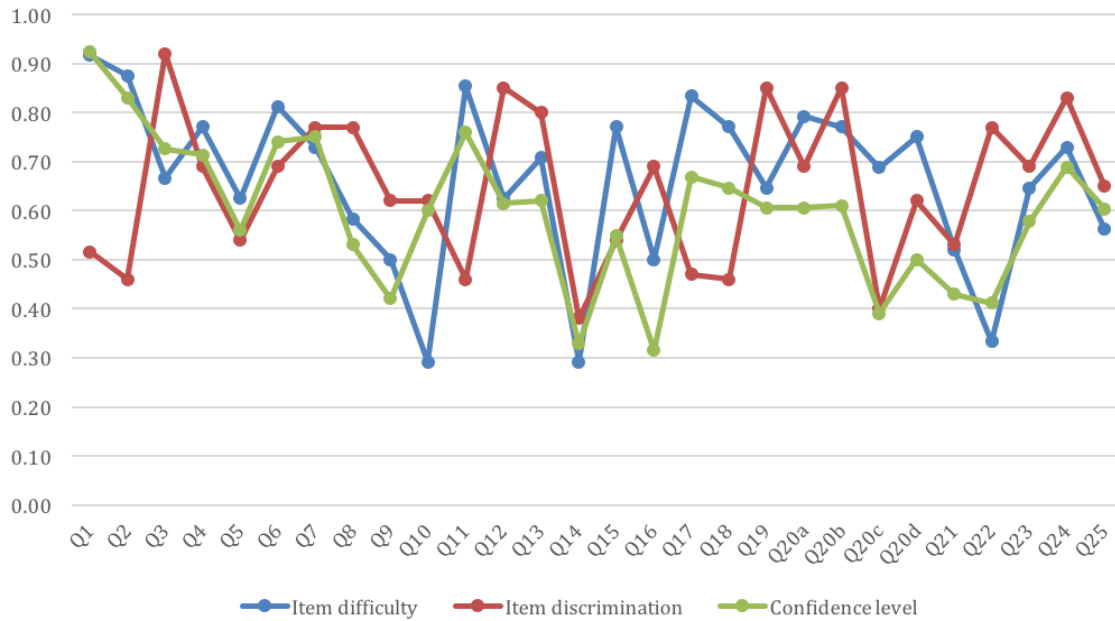


Figure 4. Items analysis on the second version of SLA for SLA survey

2.2.3 Semi-structured interviews

In order to provide a complete picture of SLA doctoral students’ statistical literacy, I supported the *SLA for SLA* survey data with follow-up, semi-structured interviews. Semi-structured interviews allowed me to probe into participants’ performance on the *SLA for SLA* survey and their experiences in using quantitative research methods. Therefore, the interview questions primarily addressed participants’ views on the survey, their general experiences with statistical analyses, and their statistical training (see Appendix D for interview questions). I interviewed 16 participants who expressed their interest in the follow-up interviews. The interviewees were from 11

SLA programs across North America. Table 3 presents detailed background information about the interviewees.

Table 3
Interviewee Data

ID	Gender	Year in Prog.	Number of Stats Courses	Departments Stats Courses Taken	Research Orientation	Interview Length
1	F	3 rd	3 (MA), 1(PhD)	Statistics	Qualitative and Quantitative	51 mins
2	F	2 nd	1(MA), 1(PhD)	Social and Behavior Sciences	Quantitative	44 mins
3	F	3 rd	1(MA), 3(PhD)	Statistics	Quantitative	37 mins
4	F	5 th	1(MA), 2(PhD)	Applied Linguistics, Educational Psychology	Qualitative	20 mins
5	F	4 th	5 (PhD)	Applied Linguistics, Math	Quantitative	34 mins
6	F	4 th	2 (PhD)	Education	Quantitative	30 mins
7	M	3 rd	2 (PhD)	Statistics	Qualitative	27 mins
8	M	4 th	2 (PhD)	Statistics, Educational Psychology	Quantitative	41 mins
9	F	1 st	1 (MA), 1 (PhD)	Education	Quantitative and Qualitative	40 mins
10	F	3 rd	2 (PhD)	Statistics, Educational Psychology	Qualitative	41 mins
11	M	3 rd	1(BA), 1(MA), 2(PhD)	Statistics, Second Language Studies	Quantitative	30 mins

Note. F = Female, M = Male

Table 3 (cont'd)

ID	Gender	Year in Prog.	Number of Stats Courses	Departments Stats Courses Taken	Research Orientation	Interview Length
12	F	2 nd	1 (PhD)	Educational Psychology	Qualitative	25 mins
13	F	3 rd	1 (PhD)	Educational Psychology	Qualitative	20 mins
14	M	4 th	1 (MA), 2 (PhD)	Linguistics, Education	Quantitative + Qualitative	35 mins
15	M	1 st	1 (BA), 1 (MA)	Linguistics, Psycholinguistics	Quantitative	25 mins
16	F	2 nd	2 (PhD)	Statistics, Applied Linguistics	Quantitative	28 mins

Note. F = Female, M = Male

2.3 Procedure

I collected the data over the course of 13 weeks. As a first step, I created an online version of the *SLA for SLA* survey via *Qualtrics* (<https://www.qualtrics.com>). *Qualtrics* is a secure, sophisticated survey software program that allows researcher to create high caliber surveys with a variety of advanced features (e.g., displaying the scores at the end of the survey). After obtaining a complete list of institutions offering doctoral degrees in SLA in North America (mostly based on Thompson, White, Loewen & Gass, 2012), I drafted a survey invitation email (see Appendix E), and forwarded it to several program directors and statistics instructors to share the link with doctoral students in their program. I also sent personal invitation emails to doctoral students whose email addresses were listed on their programs' websites in order to reach more participants. Several students noted that they also posted the link of the survey on their programs mailing list. At the end of the survey, participants completed to a second, anonymous survey where I asked them to leave their email addresses to receive a gift card and whether they would

be interested in a follow-up interview. Using this mini survey, I identified the participants who expressed their interest in participating in follow-up interviews. Then, I sent interview invitation emails (see Appendix F) to those participants to provide them with detailed information about the purpose and format of the interview. I tried to schedule the interviews usually within three days after the interviewees completed the *SLA for SLA* survey but for four interviewees, it took me more than one week to schedule an interview due to their busy schedule or late reply to the interview invitation email. I conducted all the interviews, except one, over Skype; most interviewees preferred the phone-call option on Skype. The interviews took approximately 30 minutes. To increase the rate of participation, I compensated the participants with \$10 Amazon gift cards for the survey and for the interview as well.

2.4 Quantitative Data Analysis

In the following sections, I provide details about each statistical method used in this study. I set the alpha level at .05 for all the statistical analyses used in this study.

2.4.1 Descriptive statistics

I calculated descriptive statistics on the statistical background questionnaire to examine participants' research orientation, basic statistical training, data analysis tool preferences, and self-rated statistical literacy. In addition, the results of this part were supported by the interview data.

2.4.2 Missing data analysis

Missing data is one of the most common data analysis problems, especially in survey-type studies (Tabachnick & Fidell, 2013). The missing data issue can be highly

crucial depending on the amount of missing data and the pattern of the missingness (Schafer & Graham, 2002; Tabachnick & Fidell, 2013). Missing data analysis is an important, if not necessary, step for every researcher to follow before running any statistical analyses. Indeed, Wilkinson and APA Task Force on Statistical Inference (1999) recommended that researchers analyze the missing data and report the statistical methods used to handle any missing data issues. Surprisingly, to my knowledge, this topic has received little scholarly attention in L2 research. In an effort to model good practice, I conducted a detailed missing data analysis in this study because missingness might provide further information about participants' profiles.

The first step in missing data management is to determine how much is missing. According to Tabachnick and Fidell's (2013) suggestion, if the missing data are larger than 5% in a small to moderately sized data set, the missing data issue can be serious, and thus researchers need to run further analyses. The second and probably more important step is to identify the pattern of missing data. There are three main types: a) MAR (missing at random), b) MCAR (missing completely at random), and MNAR (missing not at random). With MAR and MCAR data, there are usually no observable patterns in the missing data. That is, the missing values are randomly scattered across the data. Although MAR and MNAR data "can be problematic from a power perspective, it would not potentially bias the results" (Osborne, 2012, p. 109). Simpler and common missing data management methods such as listwise deletion can work well with such missing data sets (Scheffer, 2002; Tabachnick & Fidell, 2013). However, in MNAR data, the missing values are usually related to certain variables under study and thus "data missing not at random [*MNAR*] could potentially be a strong biasing influence" (Rubin, 1976, as cited in

Osborne, 2012, p. 109), so they cannot be ignored. More complex methods of handling missing data such as multiple imputation can produce better results in MNAR-type data sets (Scheffer, 2002; Tabachnick & Fidell, 2013).

By using the missing value analysis (MVA) on SPSS version 21, I ran a missing data analysis on the *SLA for SLA* data. As illustrated in Figure 5, the MVA results showed that 87.5% (N = 105) participants answered all the items on the survey whereas 12.5% participants (N = 15) missed some items on the survey. In total, these missing items constitute almost 7% of the survey.

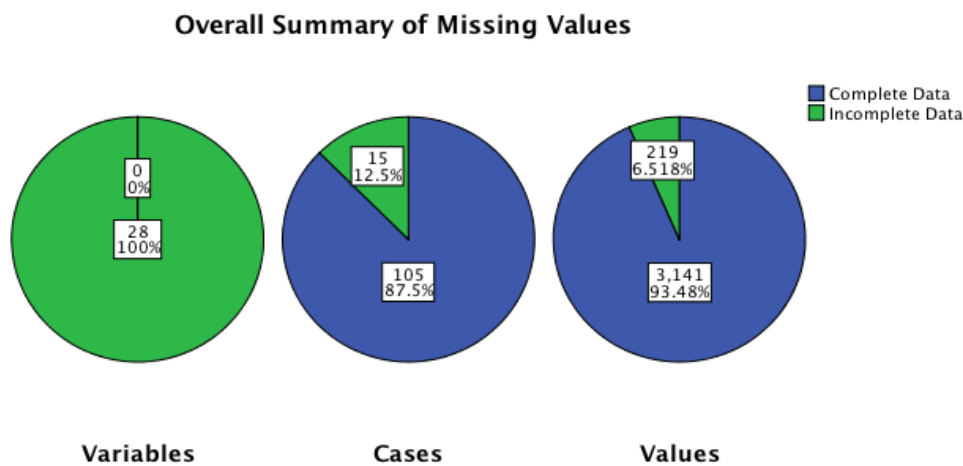


Figure 5. Missing value analysis (MVA)

Because the amount of missingness was larger than 5% (a cut-off level suggested by Tabachnick & Fidell, 2013), I decided to run a further analysis to determine the pattern of missing data and to choose a method that was the most appropriate to deal with the missing data accordingly. Based on the suggestion of Little and Rubin (2014), I first conducted Little’s MCAR test, which is essentially a chi-square test, to see whether the pattern was MCAR. The results ($\chi^2[201] = 239.437, p = .033, \text{Cramer's } V = .59$) showed

that the data was not MCAR because for a data set to be MCAR, the Little's test results should be non-significant (Little & Rubin, 2014). It was likely that the data was MNAR but there was no straight-forward statistical method like Little's MCAR test to use to decide whether the data was MNAR and missing data were related to certain variables under study. Therefore, I decided to examine six variables (i.e., considering oneself a qualitative researcher, a quantitative researcher, number of courses taken, adequacy of stats training, self-training in statistics, and self-rated statistical literacy) that I thought were potentially strong predictors of the missingness pattern in the data set. I ran several analyses (i.e., descriptive statistics, and Mann-Whitney U tests) with 27 participants with high scores (i.e., between 28 and 24) and 28 participants who did not complete all the items and got low scores (i.e., less than 13). To put it another way, I investigated whether any of the six variables listed above played a role in the missing data pattern by comparing high scoring and low scoring groups.

Mann-Whitney U test results indicated that there were statistically significant mean differences in six variables between the non-missing group and the missing group: considering oneself a quantitative researcher ($U = 100, z = -4.69, p < .001, r = -.64$), a qualitative researcher ($U = 193, z = -3.03, p = .002, r = -.41$), number of stats courses ($U = 146.50, z = -3.87, p < .001, r = -.52$), adequacy of stats training ($U = 96, z = -4.65, p < .001, r = -.63$), self-training in statistics ($U = 201, z = -2.72, p = .006, r = -.37$), and self-rated statistical literacy ($U = 89, z = -4.78, p < .001, r = -.66$). Overall, the missingness tests indicated that all of the six variables appeared to play a role in the missing data pattern. That is, the participants who were less quantitatively oriented, took few statistics courses, were not happy with the amount of their statistical training, were not doing any

self-training in statistics or had low self-rated statistical literacy score tended to not respond to some items on the survey. These results indicated that the data in this study was an example of missing not at random (MNAR) and therefore was not ignorable (Schafer & Graham, 2002).

2.4.2.1 Multiple imputation

Given that the missing data in this study were MNAR, I decided to run a multiple imputation (MI) technique on the *SLA for SLA* survey, following the suggestions of Scheffer (2002). I ran the MI on SPSS with the suggested options selected (i.e., 5 imputations and 10 iterations). After the imputation, I ended up having five different imputed data sets. Although the latest version of SPSS recognizes the imputed data file and allows researchers to automatically run many statistical tests on the aggregated imputed data file (i.e., pooled estimates), some of the advanced statistical methods such as factor analysis were not still compatible with imputed data. Therefore, I calculated the average of the five estimates for each variable imputed and created a single imputed data file for the subsequent analyses.

2.4.3 Exploratory factor analysis

Factor analysis is a series of complex structure-analyzing procedures commonly used to investigate the underlying relationship among variables in a data set (Field, 2009; Loewen & Gonulal, 2015). Exploratory factor analysis (EFA) is one of the two main types of factor analysis. As its name suggests, EFA is usually used when researchers do not have any prior expectations regarding the number of latent variables (i.e., factors or components). Further, EFA can also be used to validate a newly-designed questionnaire.

Although the content domains of the *SLA for SLA* survey were initially designed based mostly on Finney and Schraw's (2003) statistics self-efficacy survey, which measures a single factor, I decided to conduct an exploratory factor analysis on the *SLA for SLA* survey to reveal any underlying subscales of statistical literacy because this was a brand-new survey.

Before proceeding to carry out the factor analysis, I took several important factor-analytic points into account. These points comprise decisions about (a) the factorability of the data, (b) the factor extraction model used (e.g., exploratory factor analysis vs principal components analysis), (c) the factor retention criteria (e.g., Kaiser-1 rule, scree plot, parallel analysis), (d) the factor rotation methods (i.e., orthogonal vs oblique), and (e) the labelling and interpretation of the extracted factors (for a detailed review, see Loewen & Gonulal, 2015). Since particular decisions might result in distinct factor analytic results, in the following section I clearly express and tried to justify my decisions.

2.4.3.1 Factorability of the data

The first step that I took was to screen the data to see whether the data were suitable for EFA and then to check the assumptions of EFA (e.g., multicollinearity, sample size). Since EFA is based on the correlations among variables, the correlations should not be too low or too high, which indicates a lack of variability in the data. For this reason, I examined Bartlett's test of sphericity and obtained a significant result, $\chi^2(378) = 1359.446, p < .001$. This indicated that the variables were correlated and suitable for EFA. In addition to checking the correlations, having an adequate sample size is also important to have reliable factor solutions. Although, in the factor analysis

literature, there are different suggestions regarding the sample size for EFA, the minimum required sample size varies from 100 to 500. The sample size of this study ($N = 120$) met this assumption. Another, probably more reliable, method to decide whether the sample size is adequate for EFA is to check the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) value (Field, 2009). KMO values larger than 0.7 are considered good (Field, 2009). In this study, the KMO value was 0.832, which indicated a very good sample size.

2.4.3.2 Factor extraction model

There are two primary models to consider: *the component model* (i.e., principal components analysis) and *the common factor model* (i.e., exploratory factor analysis methods including maximum likelihood, principal axis factoring, etc.). However, there are two slightly different schools of thought on the differences between EFA methods and principal components analysis (PCA). One group of researchers consider EFA methods and PCA as completely different types of analyses, whereas other researchers treat PCA as a type of EFA methods (Henson & Roberts, 2006). Even though there might be theoretical differences between these two extraction models, they usually produce similar numbers of factors or components. Considering all these points, I first chose an EFA model (i.e., principal axis factoring) and then chose PCA. Although both factor solutions produced the same number of factors, the results of PCA were more interpretable when labelling the factors. In this study, I thus present the results of the PCA.

2.4.3.3 Factor retention criteria

The third step that I took was to determine the number of factors to retain. Factor analysis literature includes several suggested factor retention criteria—such as cumulative percentage of variance, Joliffe’s criterion (i.e., eigenvalues larger than 0.7), Kaiser-1 rule (i.e., eigenvalues larger than 1.0), parallel analysis and scree plot—which researchers can use to help them determine how many factors to retain. However, different criteria may lead to slightly different factor solutions (Fabrigar et al., 1999). The Kaiser-1 rule is the go-to option for researchers simple because it is the default option in many statistical packages. However, the application of EFA with only this criterion chosen tends either overestimate or underestimate the number of factors to retain (Comrey & Lee, 1992; Gorsuch, 1983). Given that, it is important to make use of more than one criterion to obtain a more reliable factor solution. Therefore, I used multiple factor retention criteria (i.e., Kaiser-1 rule, examination of scree plot, and parallel analysis) to obtain the number of factors.

2.4.3.4 Factor rotation method

After factors are extracted, these factors are rotated to produce more interpretable solutions than unrotated solutions. That is, since the first factor is usually highly loaded in a typical unrotated solution, it is suggested to rotate the factors to get better differentiation of the factors. There are two primary methods of rotations: *orthogonal rotations* and *oblique rotations*. If factors appear to be uncorrelated or independent, orthogonal rotation is suggested, whereas if factors are assumed to be correlated, oblique rotation is suggested (for a detailed review, see Loewen & Gonulal, 2015). Given that the items on the survey are correlated in nature, I decided to use an oblique rotation (i.e.,

direct oblimin) to have a better solution. In addition, I considered the items with factor loadings larger than .30 significant (Field, 2009).

2.4.3.5 Interpretation of factors

The interpretation process included examining which items loaded on which factors, and labeling each factor based on their substantive content. In this process, I paid special attention to complex variables that significantly loaded on more than one factor because complex variables make the interpretation and labeling process quite difficult. In such cases, I first attempted to consider the complex variables as an item of the factor on which the variables loaded largest. However, because one of the variables was not strictly pertinent to the content of the assigned factors, I decided to exclude this variable and reran the analysis.

2.4.4 Multiple regression analysis

I ran four multiple regression analyses by using the three factor scores and the overall survey score as outcome variables and four items on the statistical background questionnaire (i.e., quantitative research orientation, number of statistics courses taken, self-training in statistics, and year in program) as predictor variables. More specifically, I used hierarchical (also known as sequential) regression analyses by deciding the order of the predictor variables entered in the analyses (Field, 2009; Jeon, 2015).

Considering the potential impact of the predictor variables on the outcome variables, the order of entry that I used was number of statistics courses taken, quantitative research orientation, self-training in statistics, and year in program. Further, I ran additional hierarchical regression analyses in which I first entered self-training in

statistics and year in program, followed by number of statistics courses taken and quantitative research orientation to reveal which predictors emerge as significant.

Table 4

Multiple Regression Assumptions

	Minimum	Maximum	Accepted Values
Standard Residuals	-2.25	2.43	-3 to 3
Cook's Distance	.001	.058	-1 to 1
Mahalanobis Distance	.842	15.77	Below 18.47
VIF	1.02	1.69	Below 2.50
Tolerance	.59	.98	Below .40

Note. Accepted values are based on the suggestions of Allison (1999), Field (2009), and Tabachnick and Fidell (2013).

To get reliable multiple regression analysis results, I screened the data and checked the assumptions (see Table 4). First, I examined the sample size to see if the data were appropriate for regression. According to Field (2009), there should be at least 15 participants for each predictor variable. Given that, I decided the sample of 120 would be adequately large for a regression analysis with four predictor variables. Then, I conducted further data screening to see whether there were any univariate and multivariate outliers. To this end, I computed the Mahalanobis distance which is fundamentally the distance of an item from the multivariate mean (Tabachnick & Fidell, 2013). A large Mahalanobis distance indicates a potentially influential observation. However, none of the Mahalanobis distance values exceeded the critical value (i.e., $\chi^2[4] = 18.47$, $p < .001$), which was calculated based on the sample size and the number of predictors. In addition, Cook's distance, another test used to find any outliers, was within the acceptable range of -1 and 1.

Further, I checked the assumption of multicollinearity which can pose a real problem for multiple regression analysis. Thus, I examined the variance inflation factors (VIF) and tolerance values to diagnose any multicollinearity issues. Although there are no established rules of thumb, Allison (1999) suggested that if any VIF value is higher than 2.50 and the tolerance value is lower than .40, there is a reason for concern. However, there appeared to be no issue of multicollinearity in this study, with variables having lower than 2.00 VIF values and larger than .50 tolerance values. Further, I checked linearity and homoscedasticity (i.e., assumption of equal variance) by examining the scatter plots of variables and the residual plots. Overall, the results showed that the data were appropriate for multiple regression analyses.

2.5 Qualitative Data Analysis

For the qualitative part of the study, I analyzed the semi-structured interviews through a phenomenological lens. A phenomenological study describes “the common meaning for several individuals of their lived experiences of a concept or a phenomenon” (Creswell, 2013, p. 76). In relation to the purposes of the present study, this methodology enabled me to obtain a thorough description and deeper understanding of SLA graduate students’ views of statistical literacy assessment, experiences of statistical analyses as well as their background in statistical training. I followed the data analysis guidelines provided by Creswell (2013). That is, after transcribing all the interview data, I entered the data into the qualitative analysis software package, *QSR NVivo 10*. Then, I read the transcripts several times to gain a sense of familiarity of the phenomenon. After the initial readings, I tried to identify qualitatively different conceptions. Afterward, I coded these significant conceptions, exemplified by quotations, into themes and nodes.

CHAPTER 3: RESULTS

In this chapter, I present the results of the study in a question-by-question fashion. That is, I report the results separately for each research question.

3.1 Research Question 1

For the first research question, I addressed the question of the extent to which SLA graduate students have received training in quantitative research methods. Descriptive statistics obtained from the statistical background questionnaire answer the first research question. Doctoral students in the field of SLA in North America reported having taken at least two statistics courses on average ($M = 2.19$, $SD = 1.56$, $95\% CI [1.91, 2.48]$). As can be seen in Figure 6, students took statistics courses mostly in applied linguistics departments, followed by education, linguistics, and psychology departments.

On a scale from 1 (*not at all*) to 6 (*exclusively*), participants self-rated the extent to which they identified themselves as a researcher, and how frequently they conducted qualitative and quantitative research (see Table 5). It is surprising to see that the mean score for participants' considering themselves researchers is 3.71 ($SD = 1.32$). This implies that not many SLA doctoral students have embarked on conducting their own research yet. However, participants reported conducting qualitative research ($M = 3.24$, $SD = 1.36$) almost as frequently as quantitative research ($M = 3.44$, $SD = 1.44$). Indeed, the 95% confidence intervals around the means of these two items somewhat overlap, which means that it is likely that there is no statistically significant difference between

these two means at $\alpha = .05$. Figure 7 presents further information regarding the distribution of participants' responses on these two items.

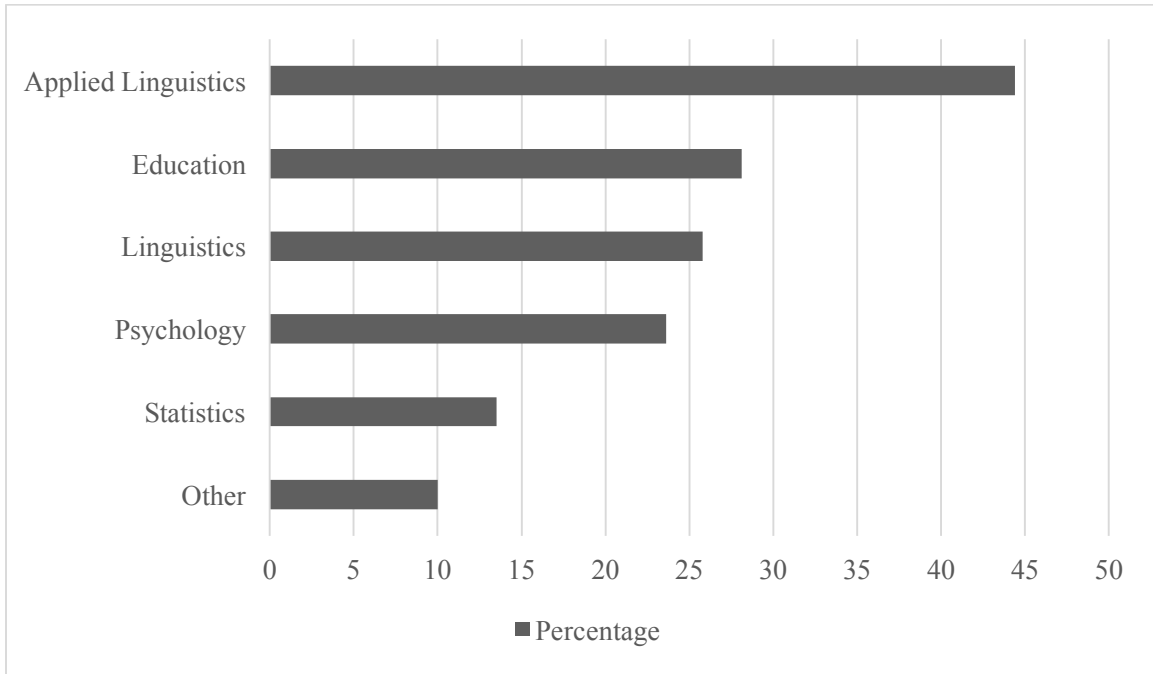


Figure 6. Departments in which statistics courses were taken

Table 5

Descriptive statistics for research orientation

	N	M	SD	95% CI
To what extent do you identify yourself as a researcher?	118	3.71	1.32	[3.47, 3.95]
To what extent do you conduct quantitative research?	116	3.44	1.44	[3.17, 3.70]
To what extent do you conduct qualitative research?	118	3.24	1.36	[2.99, 3.48]

Note. 1 = Not at all, 6 = Exclusively

In addition to descriptive statistics, I performed further analysis on participants' research orientation. I conducted a paired-samples *t* test to compare participants' self-

rated scores on qualitative and quantitative research. As indicated by the confidence intervals, there was not a statistically significant difference in their research orientation, $t(115) = 1.076, p = .284, \text{Cohen's } d = 0.14$. These results indicate that participants in this study were not biased toward those who were exclusively quantitative researcher or vice versa.

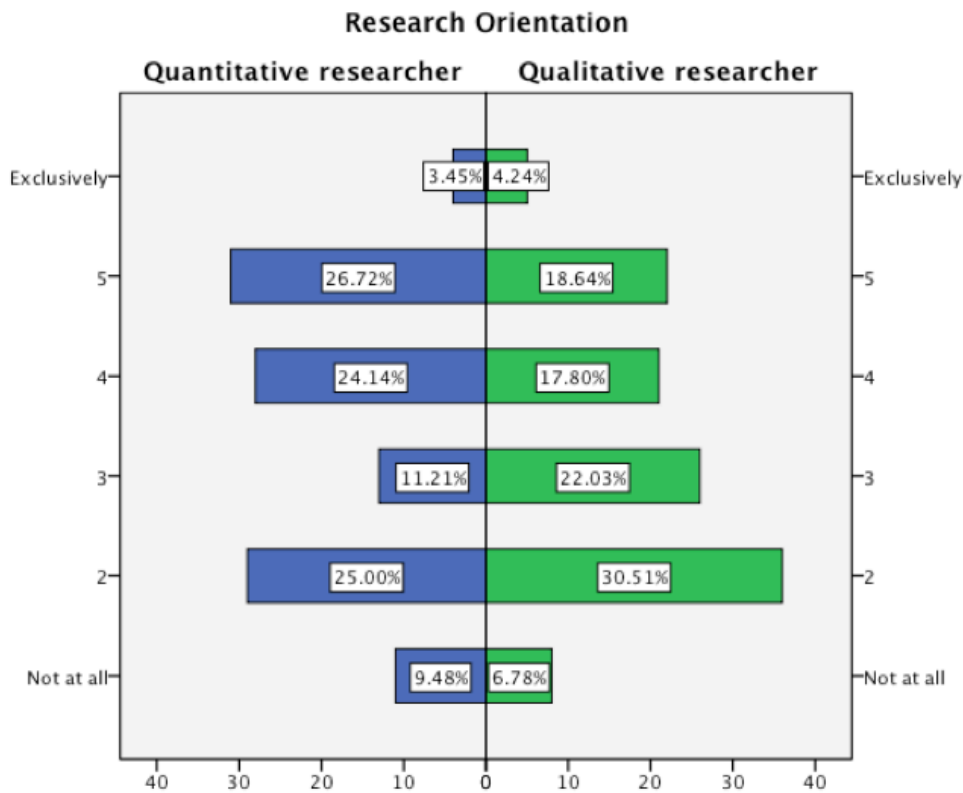


Figure 7. Participants' research orientation

Participants also rated the amount of statistical training that they have received, how satisfied they were with their statistical training, the amount of self-training in statistics and their perceived statistical literacy level, on a scale from 1 to 6. Descriptive statistics for these questions are presented in Table 6. Participants reported that they

considered themselves well-trained in basic descriptive statistics ($M = 4.58$, $SD = 1.38$, $95\% CI [4.33, 4.83]$) whereas they were less trained in overall inferential statistics ($M = 2.78$, $SD = 1.25$, $95\% CI [2.56, 3.01]$). When looking at the common inferential statistics (e.g., t test, ANOVA, chi-square and regression), the amount of training was higher ($M = 3.67$, $SD = 1.44$, $95\% CI [3.40, 3.93]$). However, as can be expected, participants had the lowest training in advanced statistics (e.g., structural equation modeling, Rasch analysis and cluster analysis) ($M = 1.91$, $SD = 1.29$, $95\% CI [1.66, 2.15]$). Due to non-overlapping confidence intervals, the difference was statistically significant between the amount of training in descriptive statistics and inferential statistics.

Table 6

Overall statistical training

	N	M	SD	95% CI
Amount of statistical training ^a				
Descriptive statistics	117	4.58	1.38	[4.33, 4.83]
Inferential statistics	116	2.78	1.25	[2.56, 3.01]
Common inferentials	117	3.67	1.44	[3.40, 3.93]
Advanced statistics	116	1.91	1.29	[1.66, 2.15]
Statistical training satisfaction ^b	116	3.20	1.29	[2.96, 3.44]
Self-statistical training ^c	117	3.00	1.41	[2.74, 3.26]
Self-rated statistical literacy ^d	117	2.90	1.25	[2.67, 3.13]

Note. ^a1 = very limited, 6 = optimal

^b1 = not satisfied at all, 6 = very satisfied

^c1 = not at all, 6 = exclusively

^d1 = beginner, 6 = expert.

In terms of adequacy of their statistical training, participants were in the middle ground. That is, they were neither dissatisfied nor satisfied with their training in statistics ($M = 3.20$, $SD = 1.29$, $95\% CI [2.99, 3.44]$). In response to the question regarding whether they do self-training in statistics, participants were again in the middle ground

($M = 3.00$, $SD = 1.41$, $95\% CI [2.74, 3.26]$). In addition, on a scale from 1 (beginner) to 6 (expert), participants rated how statistically literate they considered themselves.

Participants perceived themselves as almost average-level statistics users ($M = 2.90$, $SD = 1.25$, $95\% CI [2.67, 3.13]$).

In addition to descriptive statistics, I looked at the correlations between statistical training satisfaction, self-training in statistics and self-rated statistical literacy.

Participants' level of satisfaction in their statistical training and their self-rated statistical literacy were significantly correlated ($r = .67$, $r^2 = .45$, $p < .001$). Similarly, the amount of self-training in statistics is also significantly correlated with the level of statistical literacy ($r = .53$, $r^2 = .29$, $p < .001$).

Table 7

Type and frequency of statistical assistance

Source	N	M	SD	95% CI
Internet	118	4.27	1.54	[3.09, 4.60]
Statistical textbooks	116	3.27	1.56	[2.31, 3.69]
Colleagues	116	3.27	1.49	[2.25, 3.52]
Professional consultants	115	2.19	1.53	[1.37, 2.71]
University help center	116	1.84	1.26	[1.33, 2.52]
Stats workshop	115	1.84	1.08	[1.21, 2.10]
Other ^a	26	2.50	1.86	[1.75, 3.25]

Note. ^aAdvisor, software manuals, articles on statistics; 1 = never, 6 = very frequently

Participants reported the type and frequency of statistical help they usually sought. The most frequently reported source was the internet, followed by statistics textbooks and colleagues (see Table 7). Further, several participants ($N = 26$) also noted that they tended to consult their advisors, and read software manuals or quantitatively-oriented articles published in the field.

Table 8

Type of statistical computation

Category	<i>N</i>	%
SPSS	83	69.2
Excel	81	67.5
R	32	26.7
By hand	19	15.8
AMOS	6	7.5
SAS	5	4.2
STATA	2	1.7
I don't compute	10	8.3
Other ^a	10	8.3

Note. ^aFacets, Winsteps, Bilog, MPlus, JMP, Goldvarb, Online stats tools

Similarly, participants reported the methods by which they calculate statistics (see Table 8). SPSS and Excel were the most frequently used computation methods. The third common method was *R*. Approximately 16% of participants reported calculating statistics by hand among their preferred calculation methods.

3.2 Research Question 2

After analyzing and presenting the results of the statistical background questionnaire, I performed several statistical analyses on the *SLA for SLA* survey in order to examine how statistically literate SLA graduate students were in using statistics, which addressed Research Question 2. The average overall score on the survey was 16.38 ($SD = 7.82$, 95% $CI [14.96, 17.79]$) out of 28, which indicated that the survey was slightly difficult. The reliability of the overall survey (Cronbach's $\alpha = .891$) was quite high (Field, 2009; Kline, 1999).

Table 9

Item analysis on the SLA for SLA survey

Item	Item Difficulty	Item Discrimination	Confidence Level	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
S1Q1	.85	.38	.86	.476	.888
S4Q20	.80	.50	.64	.533	.886
S4Q21	.78	.53	.64	.618	.885
S1Q2	.75	.40	.81	.354	.889
S4Q18	.75	.53	.74	.465	.887
S4Q11	.74	.48	.78	.512	.886
S3Q17	.73	.60	.74	.564	.885
S1Q4	.73	.56	.77	.361	.889
S1Q3	.70	.63	.70	.540	.886
S2Q7	.70	.68	.71	.601	.884
S4Q23	.68	.48	.64	.455	.887
S2Q8	.67	.48	.68	.354	.889
S2Q6	.66	.53	.69	.426	.888
S5Q27	.64	.63	.74	.468	.887
S4Q22	.60	.35	.64	.348	.891
S3Q13	.58	.68	.65	.497	.886
S3Q15	.58	.33	.64	.333	.892
S4Q19	.55	.83	.72	.594	.884
S5Q28	.53	.83	.66	.616	.883
S2Q5	.53	.53	.62	.379	.889
S2Q9	.49	.55	.65	.373	.889
S4Q12	.49	.70	.68	.510	.886
S3Q24	.48	.73	.62	.537	.885
S3Q16	.43	.38	.54	.358	.891
S3Q14	.38	.53	.57	.361	.889
S5Q26	.37	.77	.62	.497	.886
S4Q25	.35	.58	.55	.387	.889
S2Q10	.32	.35	.54	.334	.892

Note. Item labels give scenario-wise information about each item. For example, S1Q1 refers to Question 1 in Scenario 1.

As shown in Table 9, I also conducted item-level analyses for the items on the survey. The table ranks the items from the easiest to the most difficult, based on item difficulty values. The smaller the item difficulty is, the more difficult an item is. According Brown (2005), items with item difficulty values below .30 are usually considered very difficult while items with item difficulty values above .70 are easy (Brown, 2005). In addition to item difficulty, item discrimination indices are in Table 9. Although majority of the items had moderate to high discrimination indices, there were a few items (e.g., S3Q15, S2Q10) with low discrimination indices close to the cut-off value (i.e., below .3) suggested by Brown (2005). As an additional analysis, I also examined confidence level scores associated with each item indicating how confident participants were in answering each item. For many items, confidence levels and item difficulty values were similar in that participants' statistical knowledge and their confidence levels were significantly correlated ($r = .78$, $r^2 = .61$, $p < .001$).

The last two columns in Table 9 are pertinent to reliability analysis. Corrected item-total correlations show how items on the survey correlate with the total score. According to Field (2009), all item-total correlations should be higher than .3 in a reliable scale. All corrected item-total correlations were above .3, which was good. Cronbach's alpha if item is deleted also provides further information about any potentially problematic items. The overall α is .891. If deletion of an item results in a substantial increase in overall alpha, then it means that particular item is problematic and thus may be dropped from the analysis. As can be seen in Table 9, although there were two items (i.e., S3Q15, S2Q10) increasing the overall reliability when deleted, the increase (i.e.,

.001) was very small. In considering all these, I kept all the items for the next statistical analysis, which is factor analysis.

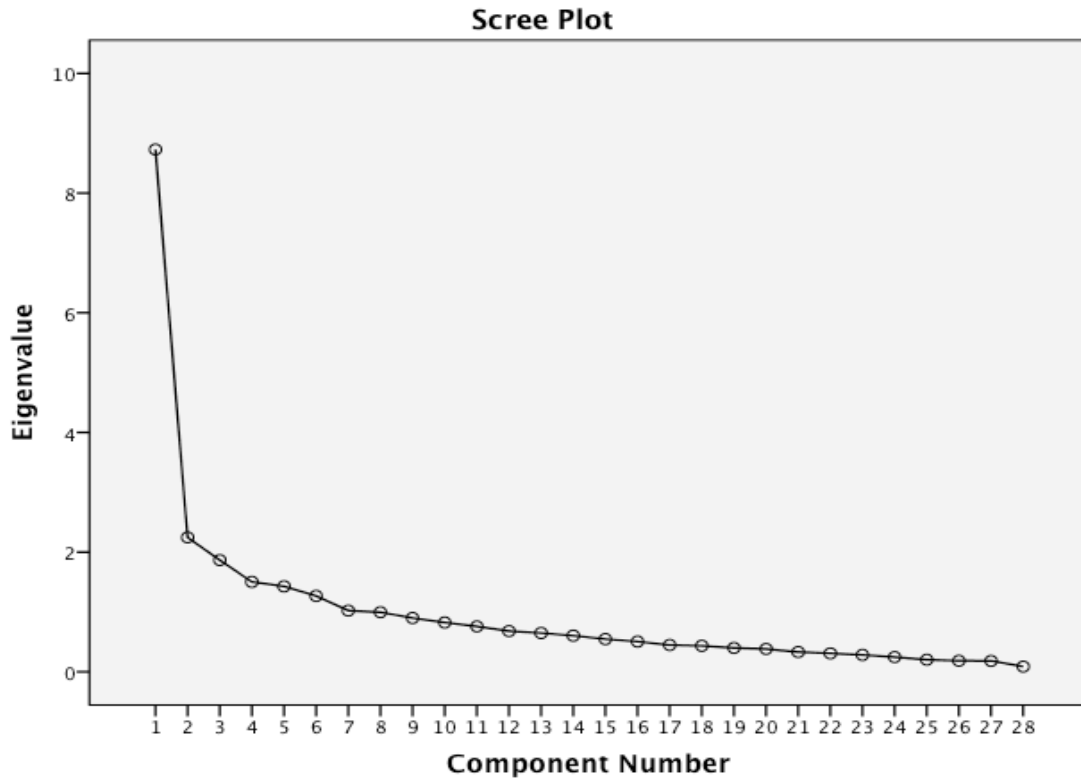


Figure 8. Scree plot for 6-component solution

I conducted an exploratory factor analysis method (i.e., principal components analysis [PCA]) to investigate any underlying constructs in the *SLA for SLA* data set, and also because it was a new survey. As discussed in the previous chapter, before running the factor analysis, I checked all the assumptions of factor analysis (e.g., from sample size to multicollinearity). The results showed that the sample size ($N = 120$) was appropriate for factor analysis ($KMO = .832$), the variables (i.e., survey questions) were correlated enough (Bartlett's test of sphericity, $\chi^2[378] = 1359.446$, $p < .001$), and there was no issue of multicollinearity (The determinant of the R -matrix was larger than

.00001). The PCA initially produced 6 factors with eigenvalues greater than 1. This six-factor solution accounted for 64.5% of the variance in the data set.

A careful investigation of the scree plot (see Figure 8) of the initial PCA analysis revealed that there were several points of inflection (i.e., components 2, 4 and 7), sharp descents in the slope of the plot. In fact, these inflection points suggested three different solutions: a one-factor solution, a three-factor solution and a six-factor solution (items before the inflection are considered in factor-solutions).

As Comrey and Lee (1992), and Gorsuch (1983) pointed out, the Kaiser's 1 rule (i.e., retaining factors with eigenvalues larger than 1.0) sometimes underestimate or overestimate the number of factors. Therefore, I used several criteria to extract a more accurate number of factors. That is, I included a parallel analysis along with the Kaiser criterion, and compared the results on a scree plot (see Figure 9). According to Hayton, Allen and Scarpello, (2004), in parallel analysis factor retention method, actual eigenvalues are compared with computer-generated eigenvalues which are created based on the same number of variables and observations as in the original data set. When the eigenvalues of the original data set are larger than parallel analysis eigenvalues, those factors are retained. Since SPSS is not compatible with parallel analysis, I used the parallel analysis engine by Patil et al. (2007) to produce parallel analysis eigenvalues. Apart from the parallel analysis criterion, I also took the cumulative percentage of variance explained by the extracted factors into consideration when deciding the number of factors to retain. As can be seen in Figure 9, the actual eigenvalues had smaller values than the parallel analysis eigenvalues starting at factor 4, which suggested a three-factor solution.

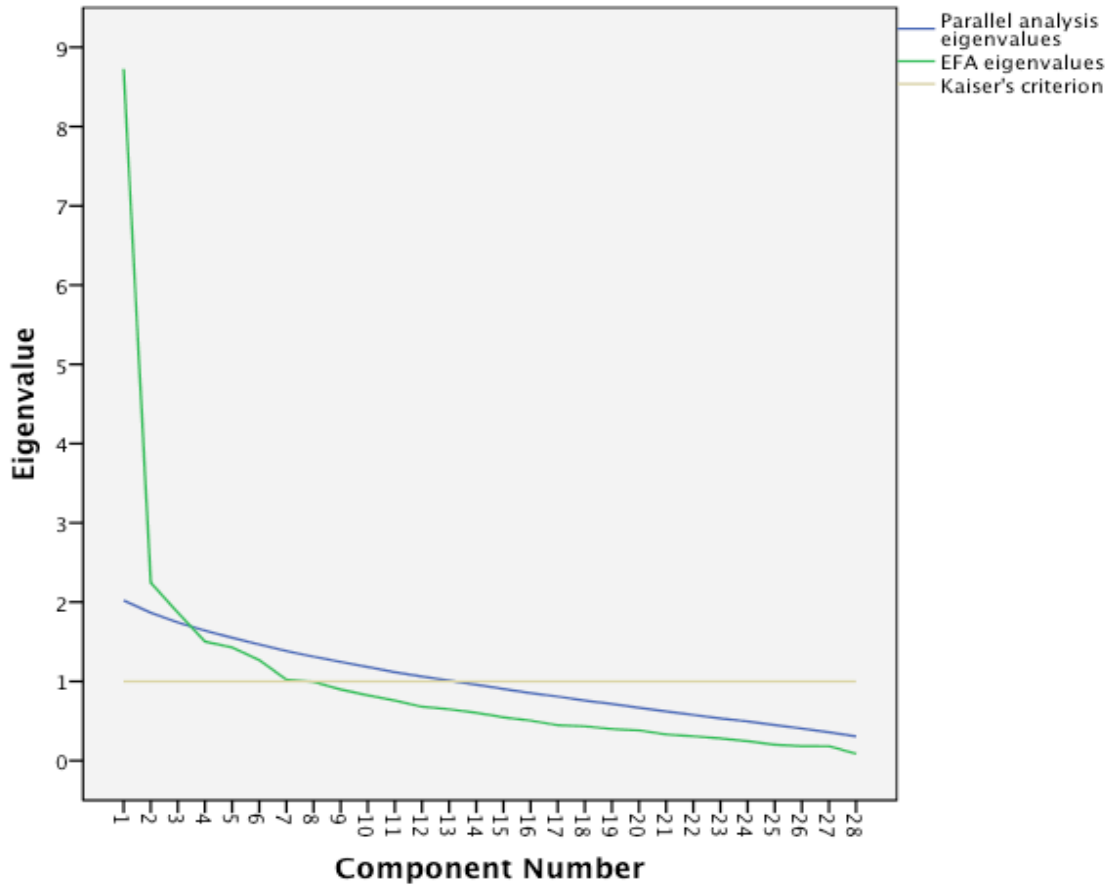


Figure 9. Visual comparison of factor retention criteria

Based on the comparison of the factor retention criteria, I decided to extract 3 factors. I reran the PCA with the 3-factor option selected. The new factor solution accounted for approximately 48% of the total variance among the variables, which was within the acceptable range (Field, 2009; Loewen & Gonulal, 2015). Table 10 presents the factor loadings for each item, and the eigenvalues, cumulative percentage of variance, and Cronbach's alpha level for each factor. I considered the factor loadings larger than .30 as significant.

Table 10

Factor loadings

Item	Factor 1	Factor 2	Factor 3
S1Q1 Understanding of sample	.708	-.015	-.081
S2Q5 Distinguishing between measures of central tendency	.695	.067	.210
S2Q6 Understanding of standard deviation	.645	.140	.091
S2Q4 Distinguishing between measures of central tendency	.520	.320	.269
S4Q20 Identifying descriptive statistics	.080	.838	.185
S4Q21 Identifying descriptive statistics	.157	.837	.230
S4Q23 Identifying inferential statistics	.156	.757	.151
S4Q18 Choosing the correct statistical test (correlation)	.078	.649	.189
S4Q22 Identifying inferential statistics	.159	.591	.293
S4Q17 Identifying type of variables	.109	.563	.393
S2Q10 Understanding of box-plot	.415	.541	.199
S1Q3 Understanding of descriptive and inferential stats	.339	.525	.360
S3Q12 Choosing the correct statistical test (chi-square)	-.181	.420	-.084
S2Q8 Identifying type of a distribution	.380	.392	-.112
S2Q9 Interpretation of box-plot	.216	.397	.380
S4Q19 Interpretation of correlation results	-.015	.195	.823
S5Q28 Interpretation of multiple regression results	.259	.071	.746
S3Q13 Interpretation of chi-square results	-.042	.208	.714
S4Q24 Understanding of type 1 error	.270	.148	.713
S2Q7 Interpretation of variance	.268	.081	.678
S5Q26 Choosing the correct statistical test (regression)	.061	.201	.605
S5Q27 Interpretation of multiple regression results	-.037	.293	.538
S3Q15 Interpretation of sample size and power	.300	.073	.500
S4Q25 Interpretation of standard error	.259	.140	.481
S3Q14 Interpretation of type II error and power	.229	.084	.469
S3Q11 Identifying type of variables	.241	.392	.402
S4Q16 Interpretation of effect size	.255	.071	.318
Eigenvalue	1.86	2.24	8.64
% of variance	6.89	8.30	31.99
Cumulative variance	6.89	15.19	47.19
Cronbach's alpha	.651	.842	.865

Note. S1Q2 was excluded from the analysis because it didn't significantly load on any factors. Also, low communality value (.118) confirmed that this item doesn't contribute to the factor solution. Shading shows factor loadings larger than .30 which were used in the interpretation of the factors.

The next step was to examine which items loaded on what factors and then to name each factor based on their main contents. Probably, the most challenging part of the

factor labeling process was to reach a decision about the complex variables, which are the items that load significantly on more than one factor. There were several instances of complex variables (e.g., S2Q4, S3Q15, S2Q8, S3Q11) in the three-factor solution presented in Table 10. Although there is no clear-cut solution to the issue of complex variables, one of the suggested solutions in the factor analytic literature is to assign the item to the factor that it loads on the highest (Field, 2009; Henson & Roberts, 2006). In some cases, it would be more reasonable to assign the item to the factor that it makes the most sense considering the overall content of the factor. For instance, it would make the interpretation of factors easier if the item S3Q11 was assigned to factor 2 instead of factor 3 because the item seemed to be more related to the items in factor 2 than those in factor 3. However, I assigned the complex variables to the factor on which they loaded most highly.

In light of these points, I labeled the first factor *understanding of descriptive statistics*, which includes items pertinent to sample, standard deviation, mean, median and mode. As for factor 2, I described it as *understanding of inferential statistics*, which contains items on correlations, chi-square, and box-plot. Although there were two seemingly unrelated items (i.e., Q20 and Q21) in this factor, I did not exclude them from the factor because these items were designed to measure participants' ability to identify whether certain statistics were descriptive or inferential. That is, the ability to label a statistic as descriptive also requires the knowledge of inferential statistics. In looking at the theme of the third factor, I considered it *interpretation of inferential statistics*, containing items that require participants to interpret the results of some common inferential statistics.

In addition to the overall reliability, I also conducted separate reliability analyses for each factor, which is a suggested procedure when a survey consists of several subscales (Field, 2009). The Cronbach's alphas for the second ($\alpha = .842$) and third ($\alpha = .865$) factors were high while the Cronbach's alpha for the first factor ($\alpha = .651$) was within the acceptable range (Field, 2009; Kline, 1999). Although the Cronbach's alpha for the first factor was slightly lower than the other factors, it is likely that this was because of the small number of items included in the first factor.

Table 11

Descriptive statistics for factors

Factors	Number of Items	M	SD	95% CI
1. Understanding of descriptive statistics	4	.73	.29	[.66, .78]
2. Understanding of inferential statistics	11	.68	.24	[.64, .73]
3. Interpretation of inferential statistics	12	.53	.27	[.49, .58]

Table 11 presents descriptive statistics for each factor along with confidence intervals. As shown in the table, the results for participants' ability to understand descriptive statistics were similar to the ability to understand inferential statistics, indicated by overlapping confidence intervals (.64 - .73 and .66 - .78). In other words, participants' success rate averaged approximately 70% on items related to both ability to understand descriptive statistics and ability to understand inferential statistics. However, participants' ability to interpret inferential statistics was significantly different from these two factors due to non-overlapping confidence intervals. That is, participants had approximately 50% success rate in answering items related to interpretation of some common inferential statistics. In fact, given that Factor 3 includes several items requiring

higher order skills (e.g., ability to interpret the results of statistics), participants' lower performance on Factor 3 is not surprising.

3.3 Research Question 3

In order to find a good model that can predict SLA graduate students' statistical literacy, which was addressed in Research Question 3, I performed four multiple regression analyses. For this purpose, I decided to use hierarchical (sequential) regression using three factors (i.e., understanding of descriptive statistics, understanding of inferential statistics and interpretation of inferential statistics) and the overall score on the survey as outcome variables and four items on the statistical background questionnaire (i.e., quantitative research orientation, number of statistics courses taken, self-training in statistics, and year in program) as predictor variables. Hierarchical regression was the better option among regression methods because in this study I looked at how different predictor variables would explain the variance in statistical literacy, while controlling for previously entered variables.

In hierarchical regression, the order of entry is often determined by theoretical or empirical importance (Field, 2009; Jeon, 2015). However, because this area of research has been relatively untapped in the field, I determined the order of the predictor variables entered in the analyses based on the potential impact of the predictor variables on the outcome variables. Thus, the order of entry was number of statistics courses taken, quantitative research orientation, self-training in statistics, and year in program. To find out whether different orders of entering would result in different results, I also entered self-training in statistics and years spent in a program first, followed by other two variables.

Table 12

Regression model summary for Factor 1

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	SEE	<i>F</i> change	<i>df</i> ₁	<i>df</i> ₂	Sig. <i>F</i> change
1	.339	.115 ^a	.107	.283	13.969	1	112	.000
2	.421	.178 ^b	.162	.274	9.004	1	111	.003
3	.429	.178 ^c	.155	.275	.031	1	110	.860
4	.429	.184 ^d	.153	.276	.616	1	109	.434

Note. ^aNumber of courses; ^bQuantitative orientation; ^cSelf-training; ^dYear in program.

First, I conducted a hierarchical multiple regression with the first factor, the ability to understand descriptive statistics, and the four statistical background items, with the order of entry as number of statistics courses taken, quantitative orientation, self-training, and years spent in an SLA program respectively.

Table 13

Model data for Factor 1

Model	B	Std.		<i>t</i>	Sig.	95%CI	
		error	β			Lower	Upper
(Constant)	.456	.083		5.05	.000	.292	.620
Number of courses	.054	.023	.237	2.377	.019	.009	.100
Quantitative orientation	.062	.025	.304	2.481	.015	.012	.112
Self-training	-.005	.023	-.025	-.224	.823	-.052	.041
Year in program	-.015	.019	-.072	-.785	.434	-.053	.023

The results in Tables 12 and 13 show that the model with all four predictors accounted for only 18.4% of the variance in Factor 1. Number of courses and quantitative orientation had significant positive regression weights, indicating participants with higher score on these variables were expected to perform better Factor 1. In fact, number of courses and quantitative orientation were the strongest predictors, accounting for 11.5%

and 6.3% of the variance respectively. However, self-training and year in program did not have any significant contribution to this model.

Table 14

Alternative regression model summary for Factor 1

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	SEE	<i>F</i> change	<i>df</i> 1	<i>df</i> 2	Sig. <i>F</i> change
1	.197	.039 ^a	.030	.289	4.542	1	112	.035
2	.204	.042 ^b	.024	.290	.322	1	111	.572
3	.369	.136 ^c	.113	.277	12.040	1	110	.001
4	.427	.182 ^d	.152	.271	6.157	1	109	.015

Note. ^aSelf-training; ^bYear in program; ^cNumber of courses; ^dQuantitative orientation.

Table 15

Alternative model data for Factor 1

Model	B	Std.			Sig.	95%CI	
		error	β	<i>t</i>		Lower	Upper
(Constant)	.456	.083		5.05	.000	.292	.620
Self-training	-.005	.023	-.025	-.224	.823	-.052	.041
Year in program	-.015	.019	-.072	-.785	.434	-.053	.023
Number of courses	.054	.023	.237	2.377	.019	.009	.100
Quantitative orientation	.062	.025	.304	2.481	.015	.012	.112

Considering that there was not prior research on this area and thus the order of entry in multiple regression analyses might make a difference, I ran alternative models where self-training and year in program were entered first. In this alternative model (see Tables 14 and 15), self-training, number of courses and quantitative orientation were significant predictors, accounting for 4%, 10% and 5% of the variance respectively. However, year in program did not have any significant contribution to this model.

Table 16

Regression model summary for Factor 2

Model	R	R ²	Adjusted R ²	SEE	F change	df1	df2	Sig. F change
1	.328	.108a	.100	.231	13.492	1	112	.000
2	.499	.249b	.236	.213	20.925	1	111	.000
3	.499	.249c	.229	.214	.021	1	110	.885
4	.503	.253d	.225	.214	.524	1	109	.471

Note. ^aNumber of courses; ^bQuantitative orientation; ^cSelf-training; ^dYear in program.

Table 17

Model data for Factor 2

Model	B	Std.		t	Sig.	95%CI	
		error	β			Lower	Upper
(Constant)	.407	.065		6.212	.000	.277	.537
Number of courses	.035	.018	.182	1.906	.059	-.001	.070
Quantitative orientation	.069	.020	.408	3.484	.001	.030	.108
Self-training	.002	.019	.011	.100	.920	-.035	.039
Year in program	-.011	.015	-.064	-.724	.471	-.041	.019

Table 18

Alternative regression model summary for Factor 2

Model	R	R ²	Adjusted R ²	SEE	F change	df1	df2	Sig. F change
1	.291	.085a	.077	.234	10.376	1	112	.002
2	.297	.088b	.072	.234	.433	1	111	.512
3	.412	.170c	.147	.224	10.765	1	110	.001
4	.503	.253d	.225	.214	12.136	1	109	.001

Note. ^aSelf-training; ^bYear in program; ^cNumber of courses; ^dQuantitative orientation.

As for the second multiple regression in which Factor 2, the ability to understand inferential statistics was the outcome variable, the model with all four predictor variables accounted for 25.3% of the variance, with number of courses contributing 10.8% and

quantitative research orientation 14.2% of the variance in Factor 2 (see Tables 16 and 17). However, self-training and year in program didn't significantly contribute the model.

Table 19

Alternative model data for Factor 2

Model	B	Std.		<i>t</i>	Sig.	95%CI	
		error	β			Lower	Upper
(Constant)	.407	.065		6.212	.000	.277	.537
Self-training	.002	.019	.011	.100	.920	-.035	.039
Year in program	-.011	.015	-.064	-.724	.471	-.041	.019
Number of courses	.035	.018	.182	1.906	.059	-.001	.070
Quantitative orientation	.069	.020	.408	3.484	.001	.030	.108

In looking at the alternative regression model (see Tables 18 and 19) where self-training and year in program went first, three variables (i.e., self-training, number of courses taken and quantitative orientation) explained around 9% of the variance whereas year in program did not fit the model.

Table 20

Regression model summary for Factor 3

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	SEE	<i>F</i> change	<i>df</i> 1	<i>df</i> 2	Sig. <i>F</i> change
1	.482	.232 ^a	.225	.235	33.836	1	112	.000
2	.640	.409 ^b	.399	.207	33.359	1	111	.000
3	.642	.413 ^c	.397	.207	.578	1	110	.449
4	.646	.417 ^d	.395	.208	.789	1	109	.376

Note. ^aNumber of courses; ^bQuantitative orientation; ^cSelf-training; ^dYear in program.

Similarly, I conducted another hierarchical regression with the four statistical background items and the third factor, the ability to interpret inferential statistics. As can be seen in Tables 20 and 21, this model accounted for 41.7% of the variance in the third

factor. The best predictor variables were number of courses and quantitative research orientation, contributing 23.2% and 17.7%, respectively. Although self-training had positive regression weights, it did not significantly contribute to the model. Likewise, year in program was not a significant predictor.

Table 21

Model data for Factor 3

Model	B	Std.		t	Sig.	95%CI	
		error	β			Lower	Upper
(Constant)	.132	.064		2.073	.041	.006	.258
Number of courses	.068	.018	.325	3.853	.000	.033	.103
Quantitative orientation	.078	.019	.419	4.050	.000	.040	.116
Self-training	.013	.018	.067	.705	.482	-.023	.048
Year in program	-.013	.015	-.069	-.888	.376	-.042	.016

Table 22

Alternative regression model summary for Factor 3

Model	R	R ²	Adjusted R ²	SEE	F change	df1	df2	Sig. F change
1	.375	.141 ^a	.133	.249	18.328	1	112	.000
2	.389	.151 ^b	.136	.248	1.372	1	111	.244
3	.574	.329 ^c	.311	.222	29.164	1	110	.000
4	.646	.417 ^d	.395	.208	16.405	1	109	.000

Note. ^aSelf-training; ^bYear in program; ^cNumber of courses; ^dQuantitative orientation.

I also found a similar pattern in the alternative multiple regression model (see Tables 22 and 23) where I changed the order of entry by entering self-training and year in program before the other two variables. This model also explained 41.7% of the variance in the third factor. In this model, number of statistics courses taken had the highest contribution (17.8%), closely followed by self-training (14.1%) and quantitative research

orientation (8.8%). However, year in program was again not a significant contributor to the model, accounting for only 0.1% of the variance.

Table 23

Alternative model data for Factor 3

Model	B	Std.		<i>t</i>	Sig.	95%CI	
		error	β			Lower	Upper
(Constant)	.132	.064		2.073	.041	.006	.258
Self-training	.013	.018	.067	.705	.482	-.023	.048
Year in program	-.013	.015	-.069	-.888	.376	-.042	.016
Number of courses	.068	.018	.325	3.853	.000	.033	.103
Quantitative orientation	.078	.019	.419	4.050	.000	.040	.116

Table 24

Regression model summary for overall score

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	SEE	<i>F</i> change	<i>df</i> 1	<i>df</i> 2	Sig. <i>F</i> change
1	.373	.139 ^a	.131	6.880	18.068	1	112	.000
2	.526	.276 ^b	.263	6.340	21.059	1	111	.000
3	.527	.278 ^c	.258	6.360	.273	1	110	.602
4	.542	.293 ^d	.267	6.320	2.347	1	109	.128

Note. ^aNumber of courses; ^bQuantitative orientation; ^cSelf-training; ^dYear in program.

Table 25

Model data for overall score

Model	B	Std.		<i>t</i>	Sig.	95%CI	
		error	β			Lower	Upper
(Constant)	8.96	1.93		4.634	.000	5.131	12.799
Number of courses	1.38	.535	.240	2.586	.011	.323	2.446
Quantitative orientation	2.34	.584	.457	4.013	.000	1.186	3.499
Self-training	-.339	.548	-.065	-.617	.538	-1.425	.748
Year in program	-.691	.451	-.131	-1.532	.128	-1.584	.203

In addition to the three components of statistical literacy, I also performed a hierarchical regression analysis considering the overall score as the outcome variable in

order to see what variables would best predict the statistical knowledge. Tables 24 and 25 present the results of this analysis. The model accounted for 29.3% of the variance. In line with the results of the previous three regression analyses, the best predictor variables were again number of courses and quantitative research orientation, explaining, respectively, 13.9% and 13.7% of the variance in overall statistical literacy score. Year in program explained only 1.5% of the variance whereas self-training did not contribute the model at all.

Table 26

Alternative regression model summary for overall score

Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	SEE	<i>F</i> change	<i>df</i> ₁	<i>df</i> ₂	Sig. <i>F</i> change
1	.252	.064 ^a	.055	7.178	7.620	1	112	.007
2	.253	.064 ^b	.047	7.209	.042	1	111	.838
3	.435	.189 ^c	.167	6.742	16.918	1	110	.000
4	.542	.293 ^d	.267	6.322	16.105	1	109	.000

Note. ^aSelf-training; ^bYear in program; ^cNumber of courses; ^dQuantitative orientation.

Similar to the other alternative regression models, three out of four variables significantly contributed the alternative model (see Tables 26 and 27). That is, number of statistics courses taken, quantitative research orientation and self-training in statistics were the best predictors, explaining 12.5%, 10.4% and 6.4% of the total variance, respectively. The only variable that did not fit the model was again year in program.

Table 27

Alternative model data for overall score

Model	B	Std.			Sig.	95%CI	
		error	β	<i>t</i>		Lower	Upper
(Constant)	8.965	1.935		4.634	.000	5.131	12.799
Self-training	-.339	.548	-.065	-.617	.538	-1.425	.748
Year in program	-.691	.451	-.131	-1.532	.128	-1.584	.203
Number of courses	1.385	.535	.240	2.586	.011	.323	2.446
Quantitative orientation	2.342	.484	.457	4.013	.000	1.186	3.499

Overall, the multiple regressions results showed that, as can be expected, SLA doctoral students who took more statistics courses, did more quantitative research, and/or did more self-training in statistics had higher scores on the statistical literacy survey.

3.4 Research Question 4

In addition to the *SLA for SLA* survey data, I conducted several semi-structured interviews to investigate SLA doctoral students' general experiences with statistics and overall satisfaction with their statistical training, addressing Research Question 4. Apart from interview data, I made use of survey takers' comments that they left at the end of the *SLA for SLA* survey and some email exchanges with participants who did not complete the survey but participated in the study through emails. I entered all the data into qualitative analysis software package, *QSR NVivo 10*, and analyzed the data through a phenomenological lens. I present the qualitative results below in a theme-by-theme fashion. Several themes emerged from the interviews and the SLA doctoral students' comments on the survey: (a) lack of deeper statistical knowledge, (b) limited number of discipline-specific statistics courses, (c) major challenges in using statistical methods, and (d) mixed-methods research culture.

3.4.1 Lack of deeper statistical knowledge

The first theme that emerged from the interview data was related to the overall content of statistics courses that participants had taken. Eight participants reported that their statistical training was mostly limited to technical know-how, with a narrow focus on the applications of statistical procedures, particularly where and when to use statistical methods. In Excerpt 1 below, Interviewee 5 reported that the statistics course that she took had a focus mostly on statistical terminologies and basic concepts.

Excerpt 1, Interviewee 5 (4th-year AL student, quantitative research orientation)

When I took the statistics course, my gut feeling was it was only about very basic concepts. So, we learned basic things like mean, median or standard deviation, something like that. The main focus was mostly on terminologies. That class was pretty fine but I really wanted to go deeper. So like, such as your survey. We need such scenarios to apply our learning, right?

Similarly, in Excerpt 2, Interviewee 7 provided a comment that although he was taught a variety of statistical concepts and procedures in his intermediate statistics course, he was clueless about when and where he could use those statistical methods in L2 research.

Excerpt 2, Interviewee 7 (3rd-year FL & ESL Ed student, qualitative research orientation)

One of the challenges I had was that we were so neck-deep in different methods of analysis like ANOVA, ANCOVA or Chi-square and all these other things. I know the names of them, but I cannot distinguish them now. And the other challenge I had was I didn't know what studies you would use them for, what studies you wouldn't use them for. I didn't understand what their shortcomings were. I didn't know when I should use one method over another method. I didn't know what type of study I could use that for.

In the same line, in the next excerpt, Interviewee 10 reported that the intermediate statistics course she took was not as in-depth as she had expected. She also added that she had still issues with choosing the appropriate statistical method for her own research.

Excerpt 3, Interviewee 10 (3rd-year SLA student, quantitative research orientation)

Even after we finished intermediate statistics course in which we covered everything like ANOVA, correlation, and regression but they were still at a basic level. So we could understand the papers we read, but we still don't know how to use like which kind of method for our own research questions.

Feelings of frustration regarding their statistics courses also echoed among the participants who completed the *SLA for SLA* survey. As can be seen in Excerpt 4 below, some of the survey takers described their statistical training as weak and felt inadequately prepared to apply statistical methods in their research.

Excerpt 4 Survey taker

The statistics course I took was like a whirlwind course, cramming everything into one semester. Therefore, I did not get a lot of hands-on, real-life research application practice. We definitely need more hands-on training in multiple statistical methods.

We often study normal samples that meet all the assumptions, and I wish we could study samples that were not normal, or did not meet all the assumptions.

Overall, participants stated that statistics courses that they had taken were too often taught with a focus on methodological technicalities. In other words, although participants might learn what certain statistical concepts and terminologies mean, they noted that they still had issues in applying their statistical skills simply because their statistical training was usually limited to technical know-how and thus lacked some other

necessary skills such as ability to use statistics properly.

3.4.2 Limited number of discipline-specific statistics courses

Although, based on the results of Research Question 1, approximately 45% of the participants reported taking statistics courses in an applied linguistics program or department, the second most prominent theme that emerged from the interview and survey data was the limited number of discipline-specific statistics courses offered by SLA programs across North America. In Excerpt 5, Interviewee 6 explicitly stated that she had to take some of the statistics courses outside her program. She also noted that because such courses were not specially designed for SLA students, the content of the courses (i.e., examples and data sets used in such courses) were not strictly related to L2 research.

Excerpt 5 Interviewee 6 (4th-year TESOL student, quantitative research orientation)

Most of the statistics courses are offered by the department of education. I think that is a big issue because if you are doing SLA, the content of the courses is a little different from, you know, SLA stuff because there are different aspects of analyzing language stuff like that. So, I think that is the biggest issue that I have faced.

In Excerpts 6 and 7, interviewees pointed out a similar issue that since their applied linguistics program could not offer most of the required statistics courses, students took those courses through different programs such as educational psychology or even statistics. However, their satisfaction with those courses was not high due to the fact that those courses were not fully addressing applied linguistics students' needs and expectations.

Excerpt 6 Interviewee 4 (5th-year AL student, qualitative research orientation)

We have to take a four course sequence quantitative research methods. The first class is within our department and then we take other courses through educational psychology department because we don't offer many in our department. And these courses were sometimes really hard to relate too our own studies. There is such a mismatch, in my opinion, between statistics classes we take and our own studies.

Excerpt 7 Interviewee 3 (3rd-year ALT student, Quantitative research orientation)

I took the course from the statistics department. So, it was not really relevant to our field and I took it in the summer so I studied with a lot of people from other departments, mostly with engineers but since I planned to minor in statistics as well so I enjoyed the course. I took a course in IRT also in the statistics department, so it was not really relevant. I mean they try to make it for education people but it is not really for language testing or applied linguistics.

Similarly, in Excerpt 8 below, Interviewee 2 noted that SLA faculty need to offer more discipline-specific quantitative research methods courses to move the field forward.

Adding to that point, the interviewee also highlighted the main reason behind the issue of limited number of statistics courses offered by SLA programs, which is the lack of qualified individuals who could teach such courses in the field.

Example 8 Interviewee 2 (2nd-year SLA student, quantitative research orientation)

I think it is a problem in our field since we are a developing field, I guess we need to offer more quantitative research methods courses, more in-house statistics courses, but the problem is do we have enough faculty who can teach such kind of courses? Well we just got a new faculty, specially hired because he has statistics background and teaches these kinds of things.

Several participants who completed the *SLA for SLA* survey also commented on the same point that although they were able to take a variety of statistics courses through different departments, they sometimes found it challenging to relate their learning to their own research.

Excerpt 9 Survey takers

Our stats classes were offered through educational psychology program because our program didn't offer them. All the examples were related to educational psychology and not applied linguistics. This is a major disadvantage. I have no idea how to apply stats to our problems. Shortly after I took these classes, our department started to host them in-house, but then stopped after one semester due to lack of funding. So, now we're in a situation where we are a highly quantitative department and really value quantitative work, but we don't even offer our own stats classes!

I am taking an intro to statistical analysis with R class right now. This is a new course offered at my department by a new professor. We were really lucky to find someone to teach a course like this, because previously we could only take statistics course from the statistics department, which was a little too advanced for most of us.

Based on the points stated in Excerpts 5 through 9, it seems that introduction to quantitative research methods courses are often offered in the field and students are then sent to outside departments for intermediate and advanced statistical training. Probably, the main reason for this is that few SLA faculty are specifically trained in teaching quantitative research methods and statistics courses.

3.4.3 Major challenges in using statistical methods

Interviewees were asked about their experience with using statistical methods in their research, along with their overall statistical training. Although interviewees articulated slightly a wide range of statistical conundrums they often faced, I present only several of these issues that featured prominently in the data. In fact, most of these issues are related, to some extent, to the first theme. The following example is a relatively common challenge that SLA graduate students tend to face when planning to use statistical methods in their research.

Excerpt 10 Interviewee 6 (4th-year TESOL student, quantitative research orientation)

The training that I received is I feel very very basic. I will be honest with you. I do not feel comfortable with a lot of things. So, if I need to do a certain test or to analyze like when I have a certain research question, I would try to reach out and ask for help. Well, basically I struggle with every element of it. Sometimes, I don't know what stats test to run or sometimes I just choose a test that I know well and use it.

Presumably, due to the lack of application-based statistical training, as reflected in the first theme in this study, statistically naïve students who were less exposed to L2 research-based statistics problems found it challenging to apply their statistics knowledge to their research. The following example illustrates this point.

Excerpt 11 Interviewee 4 (5th-year AL student, qualitative research orientation)

I am just about to be done collecting data and about to get into all my analyses. I know I am gonna have to meet my professor a lot because just kind of looking the data now, I am looking at some descriptive stats and I do survey data so I wanna look at internal validity, reliability. I am not familiar enough with it even though exactly what to put, where to get the numbers that I need. So I am gonna need a lot of refresher and a lot of help with data, I think. I feel I like I have vague ideas and I know what has to be done but I just am having hard time making the link from point A to point B.

Similarly, a survey taker commented on the same point that deciding what method would best fit their research questions was a real challenge when using statistics.

Excerpt 12 Survey Taker

I know most statistic analyses methods, but when it comes to calculating the data in SPSS, I sometimes get lost and don't really know which method I should choose for my data. I don't think I have a very clear and big picture of the whole statistics research methods and of the subtle differences between those methods.

In Excerpt 13, Interviewee 2 noted that she had issues in a slightly different stage of using statistics. That is, she described how difficult it could be to write up the results section of a quantitative study. Since the statistical software packages (e.g., SPSS) provide numerous outputs when conducting an inferential statistic, it could be challenging to know and understand when to use what output.

Excerpt 13 Interviewee 2 (2nd-year SLA student, quantitative research orientation)

I know I have a lot of difficulty in trying to explain the results in writing. I mean more or less I can understand and interpret tests like ANOVA, multiple regression but to put it into writing sometimes is difficult. Even though I was taught what to report like f-value, degrees of freedom, I am not sure if the way I report is correct or if I need to report every single time. I think those are the issues that I face when I use statistics.

Closely related to the point stated in Excerpt 13, several interviewees noted that they had issues in deciding what to report and what not to report, apart from carrying out statistical analyses. Indeed, considering that SLA is a young but developing field, the field needs clear, field-specific standards for reporting practices. Although there are a few widely-accepted guidelines such as the APA manual in the field, it seems SLA students try to look for easier ways to report statistics. Excerpt 14 clearly illustrates this point.

Excerpt 14 Interviewee 3 (3rd-year ALT student, Quantitative research orientation)

I don't have official or specific guidelines I think. Basically I try to follow APA and manuals. Sometimes, it takes a while to find where the information is in manuals because they don't seem to have a lot of information about how to present numbers, different, new analyses. So, I try to look at other articles in the field, in my field to see how they report things. So, sometimes I just try to find some well-known researchers in my field and follow the way they report.

In some cases, reporting practices seem to be more related to the statistical literacy levels of participants rather than the purpose of information transparency and richness (see

Excerpt 15 below). In other words, being fully capable of performing a statistical test and then deciding what should get reported is indeed an important part of statistical literacy.

Excerpt 15 Interviewee 10 (3rd-year SLA student, quantitative research orientation)

When we took the course exams, we were shown very long computer output from descriptive information through like everything but when it comes to our own research, it is sometimes hard what to report and what to exclude.

In addition to reflecting on their own experiences with using statistics, several interviewees also discussed their perceptions of the statistical knowledge of graduate students in the field. Although the use of statistics has increased over the years, the methodological quality in L2 research seems to be less than optimal. In other words, how well L2 researchers adhere to standards of methodological rigor when carrying out certain statistical methods is still not at a desired level. In Excerpt 16, Interviewee 2 stated that most SLA graduate students have problems with using and interpreting statistical analyses, and consequently depend on the default options in statistical software packages when performing certain statistical methods.

Excerpt 16 Interviewee 2 (2nd-year SLA student, quantitative research orientation)

Honestly, I feel like most people kind of well at least grad students-wise, I think they just use SPSS and look for things that look right like they know they are supposed to do kind of analyses so they just rely on SPSS to just do it for them but without really understanding what they are doing and why they are doing.

Also related to the above point, there might be differences between what L2 researchers really know about statistics and how they use statistics in their research, as illustrated in Excerpt 17 below.

Excerpt 17 Interviewee 6 (4th-year TESOL student, quantitative research orientation)

Based on my observations, some researchers try to avoid stats or they invite somebody else who has the expertise. They are like “Oh I don't mind putting this person as a second author if they do my stats for me.” It is very common notion I keep hearing. Similarly, you see students who are not that good at stats but when they publish they have superior stats in their paper. Obviously, they are getting help from somebody. So, it is very hard to tell because people use different resources.

3.4.4 Mixed-methods research culture

As several methodological reviews (e.g., Gass, 2009; Lazaraton, 2000, 2005; Norris, Ross & Schoonen, 2015; Plonsky, 2011, 2013, 2015) highlighted, quantitative research methods predominate L2 research. In line with this, L2 researchers usually consider themselves either as a qualitative researcher or a quantitative researcher. In such cases (see Excerpt 18), strict research orientation can influence researchers' willingness to expand their knowledge of other research methods.

Excerpt 18 Email Exchange

As emerging scholars, I think we should all strive to become more knowledgeable on any tools that can help us answer or develop our research questions (regardless of their methodological or epistemological orientations/implications), which is why I begin by pointing out to how useful your survey was in highlighting my illiteracy in stats. Your survey made it clear to me that I could definitely use a statistics course to enrich my researcher skills and consider some qualitative + quantitative tools in the future. I also think that my qualitative bias as an emerging scholar trying to position myself as a qualitative researcher has contributed to my lack of stats literacy.

Apart from these two paradigms of research, there is also mixed-methods research that can serve as a bridge between qualitative and quantitative research. Although there are two dominant research cultures in L2 research, it seems a third research culture is also

slowly emerging. Indeed, as can be seen in Excerpts 19 and 20 below, while interviewees noted that there were some researchers who were at the extreme ends of the qualitative-quantitative dichotomy, they were glad to see more L2 researchers were adopting an eclectic method instead of a mono-paradigm approach, which can result in superior research.

Excerpt 19 Interviewee 1 (3rd-year SLA student, quantitative and qualitative research orientation)

I think there is a huge disconnection between qualitative and quantitative analyses. That is really hard to overcome. Because I was strictly thinking quantitative analysis in my master's thesis. I would have used a mixed methods approach if I had had that perspective, mixed method perspective in advance rather later. I have seen students in my program like students either like stats or hate stats. There is usually no middle ground. So in that regard I am an outlier because I think like stats methods are super cool even though I am not going to use them for my dissertation. Also I think the number of researchers who conduct mixed methods research is increasing recently. I know there is a professor who encourages her students to do mixed methods studies but I think there are not many people who have both perspectives.

Excerpt 20 Interviewee 7 (3rd-year FL&ESL student, qualitative research orientation)

I also took a course here that falls under qualitative but it was like a mixed-methods course which I really enjoyed because I am sure you realized that for most people, there is a dichotomy. They are either strictly quantitatively-oriented or strictly qualitatively-oriented. Even though the statistics is hard for me, I really appreciate it. That is why I like mixed-methods because you can implement them. I am glad that mixed-methods approach is getting more exposure and more respect.

Overall, several themes emerged from the interview data regarding SLA doctoral students' experiences with using statistics and their statistical training. First, a number of interviewees pointed out that the statistical training that they received in their programs was too often limited to statistical terminologies and concepts. Several interviewees,

however, expressed that they need deeper statistical knowledge to deal with the complex phenomenon of L2 research. Second, it appears that discipline-specific statistics courses, particularly intermediate and advanced statistics courses, are not common in the field of SLA. Although approximately half of the participants reported taking a statistics course in their own program, they also called for the need for more in-house statistics courses in which the examples and data sets used are more applicable to second language research. Third and probably mattering most is related to the challenge that SLA doctoral students often encounter when using statistics in their research. The qualitative data revealed that doctoral students had issues in almost every aspect of applying statistics, from choosing the most appropriate statistical method for their research questions to deciding what and how to report. Finally, mixed-methods research as an emerging paradigm in the field of SLA has been acknowledged by several interviewees.

CHAPTER 4: DISCUSSION

This study is novel in the field of SLA in that to date, no study has been conducted to directly measure the statistical knowledge of SLA doctoral students. Moreover, the secondary purpose of this study was to provide a snapshot of SLA doctoral students' training in statistics and experiences with using statistics. Therefore, the results of this study will provide new insights as to the status of statistical literacy in the field, through the lens of doctoral students, who are an important element of SLA programs.

In the following sections, I discuss the results of the study in depth in light of the statistical literacy studies conducted in other neighboring fields such as psychology and education. I provide a result-by-result discussion in this chapter. That is, I first interpret and discuss the results of the first research question addressing the extent to which doctoral students in the field of SLA in North America have received statistical training. Second, I address the results of the second research question pertinent to how statistically literate SLA doctoral students were. Next, I address the results related to what variables play a key role in statistical knowledge of the doctoral students in the field. In addition, I provide a detailed discussion of the results obtained from the qualitative data, by drawing on the results of the other research questions whenever possible. Finally, I discuss the limitations, and conclude the chapter with several suggestions for SLA graduate students, statisticians³, and SLA programs.

4.1 Statistical Training in SLA

The first research question broadly dealt with the status of statistical training among doctoral students in the field of SLA, focusing on various aspects of

methodological training such as number of statistics courses taken, research orientation, type and frequency of statistical assistance and computation, statistical training satisfaction, self-training in statistics and perceived statistical literacy. The results indicated that the average SLA doctoral students had taken at least two statistics courses ($M = 2.19$, $SD = 1.56$). In addition, approximately 45% had taken statistics courses in applied linguistics programs or departments. These results to some extent echo the findings of other similar studies in the field (i.e., Gonulal et al., in preparation; Lazaraton et al., 1987; Loewen et al., 2014). In their pioneering study looking at applied linguists' literacy in statistics and research methods, Lazaraton et al., (1987) reported that applied linguists took two research methods courses (including both qualitative and quantitative research methods) on average ($M = 2.27$, $SD = 2.18$). Loewen et al.'s (2014) partial replication of Lazaraton et al.'s survey showed that doctoral students had taken approximately two statistics courses ($M = 1.88$, $SD = 1.78$) and roughly 30% of these courses were taken in applied linguistics and SLA departments. It appears that the field has made some progress in regards to the number of statistics courses taken over 2.5 decades. Indeed, in a more recent study looking at the statistical literacy development of SLA graduate students (i.e., both MA and Ph.D. students), Gonulal et al., (in preparation) also found a similar number of statistics courses reported ($M = 1.75$, $SD = 1.35$) and almost one-fourth of participants had taken a statistics course in applied linguistics departments.

When compared to the findings of these three discipline-specific studies, the results of this study indicate a non-negligible increase in statistical training in the field of SLA in North America, although there might be some participant-wise overlap with

Loewen et al. and Gonulal et al. In fact, given that the sample of this study consisted of roughly similar numbers of qualitatively-oriented and quantitatively-oriented students, this increase in statistical training appears to be more significant. However, this finding is still noticeably different from the amount of statistical training in sister disciplines. For instance, the average number of statistics courses required in education doctoral programs is 3.67 ($SD = 1.91$) (Leech & Goodwin, 2008) whereas the average time to complete graduate level statistics courses in psychology is 1.2 years (Aiken et al., 2008). Although the field of SLA seems to be still behind other neighboring disciplines in terms of statistical training, the slight increase in the number of statistics courses taken along with the increased percentage of statistics courses taken in SLA programs provides a reason to be optimistic about the future of statistical training in the field in North America.

Of course, the number of statistics courses taken does not necessarily ensure higher level statistical knowledge. The content of the statistical training is also equally important. When looking at the amount of the statistical training that SLA doctoral students received in three distinct areas of statistics (i.e., basic descriptive statistics, common inferential statistics and advanced statistics, as grouped by Loewen et al., 2014), as might be expected, SLA doctoral students considered themselves well trained in descriptive statistics ($M = 4.58$, $SD = 1.38$) including concepts and procedures such as mean, median and standard deviation. However, their self-rated training in inferential statistics ($M = 2.78$, $SD = 1.25$) is significantly lower. In particular, participants reported they had the lowest training in advanced statistics ($M = 1.91$, $SD = 1.29$). Perhaps, a direct interpretation of these results might be that the majority of the statistics courses taken by SLA doctoral students focused mostly on basic statistics and partially, on

intermediate statistics. It seems that SLA doctoral students are rarely taught advanced statistics. Although this situation is not completely different in other disciplines (e.g., counseling, education, and psychology) where more extensive training in advanced statistics is suggested, if not required (Aiken et al., 2008; Borders et al., 2014; Leech & Haug, 2015; Rossen & Oakland, 2008), specialty statistics courses such as a full-semester course on regression, ANOVA or structural equation modelling, which can provide thorough training in certain statistical procedures, are at least more common than in the field of SLA.

To put it briefly, the overall statistical training in the field seems to be limited to largely introductory, and partially intermediate concepts and procedures. Indeed, regarding the adequacy of their statistical training, SLA doctoral students were moderately satisfied with their training in statistics ($M = 3.20$, $SD = 1.29$). It is also reflected in the interviews that interviewees felt their training was mostly inadequate. This finding is largely consistent with Loewen et al.'s study, in which 47% of doctoral students felt that their statistical training was somewhat adequate, 40% felt that their training was inadequate while only 13% was happy with their training.

It is important to note here that taking statistics courses is not the only source of gaining and improving knowledge in statistical methods. It is quite possible that student might improve their statistical knowledge outside of the classroom. Especially given that SLA doctoral students reported frequently using the Internet and statistical textbooks for statistical assistance, one might think that they can develop and expand their knowledge in statistics in a self-taught manner. Unfortunately, this study suggested otherwise. Self-training in statistics was not very common among SLA graduate students ($M = 3.00$, SD

= 1.41). This finding somewhat aligns with Golinski and Cribbie's (2009) claim that not many graduate students in psychology programs tend to improve their knowledge of statistical methods through self-training.

4.2 Statistical Literacy in SLA

After providing a contemporary picture of the state of statistical training in the field of SLA in North America, I now turn to the question of how statistically literate SLA doctoral students were. While there has been some interest in SLA researchers' training in quantitative research methods (Gonulal et al., in preparation; Lazaraton et al., 1987; Loewen et al., 2014), there has been a lack of instruments that can accurately assess SLA researchers' knowledge of quantitative research methods. Given this lack, I and a group of SLA researchers with reasonable knowledge in statistics developed a discipline-specific statistical literacy survey based on Finney and Schraw's (2003) statistics self-efficacy survey (see Chapter 2 for further details about the instrument development process). I attempted to measure doctoral students' knowledge of statistics through this instrument. Before moving on to how knowledgeable they were in statistics, I briefly discuss the components of this survey.

When looking at the factor structure of this survey, principal components analysis revealed three components of statistical literacy: a) understanding of descriptive statistics, b) understanding of inferential statistics, and c) interpretation of inferential statistics. This finding mostly corroborates previous studies on statistical knowledge. Although it may not be completely, in a factor-analytic study dealing with the teaching of statistics in statistics departments, Huberty et al. (1993) also identified three domains of statistical knowledge including procedural knowledge, knowledge of simple concepts and terms

related to statistics, and conceptual understanding (linking two or more statistical concepts and procedures).

In looking at statistical literacy studies, the three-component statistical literacy found in this study is largely consistent with Watson's (1997) three-tiered model of statistical literacy. Watson developed her model based on the models of learning from developmental psychology. In her model, the first tier includes a basic understanding of statistical concepts such as percentage, median, mean, odds, probabilities and measures of spread. Building on the first tier, the second tier includes understanding of commonly encountered statistical concepts in a social context. The third tier, the highest level in Watson's model of statistical literacy, includes questioning statistical conclusions and results. Watson noted that the skills used in the third tier represent higher-order thinking. Indeed, the third component of the statistical literacy in this study also appeared to pertain to a more sophisticated way of thinking. Additionally, the groupings in this study, to some extent, overlap with the five statistical knowledge elements of Gal (2002). These five elements are: "a) knowing why data are needed and how data can be produced, b) familiarity with basic terms and ideas related to descriptive statistics, c) familiarity with graphical and tabular displays and their interpretation, d) understanding of basic notions of probability, and e) knowing how statistical conclusions or inferences are reached" (p. 11).

Related to the notion of statistical literacy, Schield (2010) made a distinction between statistical literacy and statistical competence, adding that the former is needed by students in non-quantitative majors such as English, education, history that "have no quantitative requirements" whereas the latter is needed by students in quantitative majors

such as economics, biology, and psychology that “have a statistics requirement” (p. 135). His definition of statistical literacy includes “the ability to read and interpret summary statistics in the everyday media: in graphs, tables, statements and essays” whereas his definition of statistical competence comprises “the ability to produce, analyze and summarize detailed statistics in surveys and studies” (p. 135). It seems that based on his definitions, data consumers need statistical literacy while data producers need statistical competence. However, I look at the concept of statistical literacy from a broader aspect and thus I believe that SLA researchers—although it may not be fair to consider all SLA doctoral students as future academics— need both statistical literacy and statistical competence as consumers and producers of L2 research.

In the same way, Ben-Zvi and Garfield (2004) (and also Garfield & Ben-Zvi, 2007) used slightly different terms regarding statistical literacy. Specifically, they highlighted the distinctions between statistical literacy, statistical reasoning, and statistical thinking. Based on all these different definitions and descriptions, statistical literacy includes the ability to know and understand basic statistical terms; statistical reasoning is more related to the ability to interpret statistical information and statistical results; and statistical thinking involves the knowing how and why to use, for example, a certain statistical method, and also the ability to critique and evaluate the results of a statistical study. Considering these points, it seems that the first two components of the *SLA for SLA* survey align with Ben-Zvi and Garfield’s (2004) statistical literacy definition whereas the third component appears to be more related to statistical reasoning.

In considering more SLA-oriented research, these results are mostly in line with the categories of self-rated knowledge of statistical concepts in Loewen et al.’s (2014)

statistical literacy study. Loewen et al. also found three categories of statistical knowledge: a) basic descriptive statistics knowledge, b) common inferential statistics knowledge, and c) advanced statistics knowledge.

When viewed in its entirety, it seems that two elements of statistical knowledge are somewhat common across all these studies: knowledge of descriptive statistics and knowledge of more sophisticated statistical methods, which could be broadly considered inferential statistics. Similarly, although statistical literacy, as highlighted in previous studies (Ben-Zvi & Garfield, 2004; Gal, 2002, 2004; Garfield & Ben-Zvi, 2007; Schield, 2010; Watson, 2002; and others) appeared to comprise a variety of interrelated skills, the results of the factor analysis in the present study indicated that statistical literacy can be broadly categorized as the ability to understand and use statistical concepts, and the ability to interpret and critically evaluate statistical information represented in tabular and graphical forms.

Returning to the question of how statistically literate SLA doctoral students were, the results of this study revealed that overall, SLA students were good at understanding both descriptive statistics (i.e., Factor 1) and inferential statistics (i.e., Factor 2) whereas their performance on interpreting inferential statistics (i.e., Factor 3) was significantly lower. More specifically, SLA students were able to answer the items testing the knowledge of mean, median, standard deviation, t test, ANOVA, chi-square, and correlation, with an approximately 70% success rate, but when it came to the items requiring not only some knowledge but also interpretation of statistical procedures such as chi-square, correlation, and regression, they were able to find the correct answers, with an approximately 50% success rate. Given the fact that the third factor (i.e., interpretation

of inferential statistics), much similar to Ben-Zvi and Garfield's (2004) statistical reasoning, consists of several items requiring higher order skills and more sophisticated knowledge of statistical concepts, it is presumably not unexpected that SLA graduate students could not perform as well as they did on the other two factors.

To my knowledge, there is no published research with which to compare the results of this study directly. However, several quantitatively-minded researchers have raised some concerns in regards to the use of certain statistical tests by L2 researchers. For instance, Norris (2015) highlighted a number of issues associated with the use and interpretation of significance tests in the field of SLA. Likewise, although it may be tangentially related to the issue of using and interpreting statistics, Plonsky (2011, 2013, 2014), in his extensive work looking at methodological quality of quantitative L2 research, also argued that although some inferential statistics such as *t* test, ANOVA, chi-square, correlations and regressions, along with descriptive statistics, are commonly used in SLA research, there are some issues regarding reporting of inferential statistics. More specifically, reporting of *F*, *t*, *p* and chi-square values, means, standard deviations, confidence intervals, and effect sizes is "far from perfect" (Plonsky, 2014, p. 458).

Overall, it seems that SLA doctoral students possess fundamental knowledge of basic and common inferential statistics but when it comes to interpretation of statistical information, their skills are still less than optimal. Among many possible reasons, this current state of statistical literacy among SLA doctoral students might to some extent be attributed to the amount and quality of statistical training they received, particularly, in common inferential and advanced statistics. Another reason might pertain to the

frequency of self-statistical training. In the following section, I discuss some of these potential factors.

4.3 Predictors of Statistical Literacy

Another purpose of this study was to investigate what variables would be predictive of statistical literacy. Presumably, many L2 researchers would suggest that number of courses taken in statistics alone is predictive of statistical literacy. Although a few studies (e.g., Gonulal et al., in preparation; Loewen et al., 2014) have examined what variables play a role in L2 researchers attitudes towards statistics and statistical self-efficacy, many questions still remain in this area.

The results of the multiple regression analyses revealed that number of statistics courses taken, quantitative orientation, and self-training in statistics were significant predictors of not only the individual components of statistical literacy but also statistical literacy as a whole. These findings suggest that as expected, SLA students who took more courses in statistics, did more self-training in statistics or did more quantitative research tended to have higher statistical knowledge. When looking at other similar studies, similar findings have been reported. For instance, in his study using path analysis to develop a model to explain statistics achievement among graduate students in social and behavioral sciences, Onwuegbuzie (2003) found that number of college-level statistics courses taken were negatively correlated with statistics anxiety but positively correlated with statistics achievement. Similarly, Estrada, Batanero and Lancaster (2011) also found number of statistics courses taken to be positively affecting statistical knowledge and attitudes towards statistics.

As for L2-oriented research, Loewen et al. (2014) found a similar result in that number of statistics courses an individual took was a significant predictor of attitudes towards statistics and statistical self-efficacy. In their study examining the development of statistical literacy among SLA graduate students during semester-long statistics courses, Gonulal et al. (in preparation) indicated that SLA students made significant gains in their ability to interpret and use inferential statistics. Further, they also found significant gains in students' statistical self-efficacy. In addition, several studies in education (Capraro & Thompson, 2008; Henson et al., 2010) and psychology (Aiken et al., 2008; Golinski & Cribbie, 2009; Rossen & Oakland, 2008) have anecdotally reported that number of statistics courses plays an important role in graduate students' statistical knowledge development. Overall, all these studies collectively suggest that statistics courses are crucial elements of statistical literacy.

Quantitative research orientation was also a significant factor in statistical literacy. This means that SLA doctoral students with a stronger quantitative orientation appeared to have better knowledge of statistical analyses. It is well known that there are two main types of research methodology dominating the field of SLA, but a third one (i.e., mixed-methods approach) is also slowly finding its way into the field (I will discuss this in detail later in this chapter). These two camps of research methodology have unique and complementary advantages, and thus require different sets of skills and challenges on the part of the researchers (Creswell & Clark, 2011). Therefore, an individual's research orientation (i.e., qualitative and quantitative) obviously affects their development as a researcher, or vice versa. In other words, researchers who embrace a more quantitative research orientation would probably want to improve themselves in areas related to

quantitative research methods, and engage in more quantitatively-oriented research. That is, it is highly likely that quantitatively-oriented students tend to take more statistics courses and do self-training more frequently. In looking at L2-specific studies, this finding is consistent with Loewen et al.'s (2014) study in which quantitative orientation was found to be a strong predictor of statistics self-efficacy whereas qualitative orientation did not significantly contribute to statistics self-efficacy scores.

Aside from the above-discussed factors influencing statistical literacy, alternative multiple regression analyses also indicated that self-training in statistics had a statistically significant impact on the statistical knowledge scores. Although, as alluded to earlier in this discussion chapter, self-training is relatively infrequent among SLA graduate students, it is gratifying to see that self-training is an important contributor of statistical literacy. However, year spent in program towards a doctoral degree was not a significant predictor of statistical literacy, especially considering that doctoral students in the field are likely to gradually engage more in conducting research (e.g., qualifying research paper, dissertation) towards the end of their graduate education. A strong interpretation of this finding would be that since most SLA doctoral students are often done with course work within two years after entering the SLA program (Thomas, 2013), students probably stop taking quantitative research methods courses after that, unless they have a special interest in certain statistical methods that they plan to use in their own research, or have a quantitative research orientation, or do more self-training. It is also probable that any variance accounted for by years spent in a SLA program might be subsumed by courses and/or orientation. However, all these are speculative. Thus, further research is certainly needed in this area.

4.4 A Glimpse into Pandora's Box: Issues Related to Statistical Training and Using Statistics

The last research questions asked in this study focused on SLA doctoral students' overall satisfaction of their statistical training and experiences with using statistics. Results from the analysis of the interview data and the comments left at the end of the *SLA for SLA* survey provide a snapshot of the current state of statistical training and statistical literacy, in particular looking at the issues that are common among SLA doctoral students in North America. In fact, findings here are mostly in line with the results of the previous research questions addressed in this study.

First, the interviewees pointed out several issues regarding the content and format of the statistics courses that they had taken, especially in non-SLA departments. More specifically, several interviewees noted that some statistics courses tended to lack the necessary breadth and content, and were often limited to methodological technicalities, with a narrow focus on reasoning. Probably, the main issue pertains to the limited hands-on experience opportunities offered in those courses because research skills can, to a great extent, be acquired by doing. In looking at the literature on teaching statistics to non-statistics majors, Yilmaz's (1996) real-data approach which involves a good proportion of hands-on activities, along with relating statistics to real world problems (for a review, see Brown, 2013).

Further, interviewees reported that some statistics courses they had taken did not have enough L2-specific content to equip SLA students with the necessary knowledge to employ statistical analyses within L2 research. In fact, this finding might be a direct consequence of the inadequate number of discipline-specific, particularly higher level,

statistics courses offered by SLA programs. Consequently, some SLA programs send their students to other departments, for instance, for intermediate and advanced statistics. The problem here appears to lie in the fact that these interdisciplinary courses are not necessarily designed to address the statistical methods that can allow investigating the complex nature of L2 research. In more concrete terms, there appears to be a difference between the examples and data sets used in such courses and in the courses offered in the field. Therefore, although statistical procedures offered in any departments are, and should be, theoretically and conceptually, the same, different issues may arise in application. To illustrate, small sample size (generally less than 20, Plonsky, 2013) in L2 research can be a real issue as it creates a problem for statistical power, whereas it may not be that much of a problem in other disciplines. Because in order to have a complete picture of how second languages are learned, one needs to go beyond overstudied languages (e.g., English, Spanish, German) and focus on linguistic features that are unique to understudied languages. Therefore, as Plonsky (2011) stated “it may not be fair to hold SLA to the same standard or expectation of large samples as one might in a field such as psychology where researchers often have access to undergraduate participant pools or otherwise larger populations” (p. 83). However, it is still crucial to take courses in neighboring fields to broaden our knowledge of available statistical procedures, even if statistical methods learned in other fields may not always be easily applied to L2 research. Considering all these points, L2 researchers should be more knowledgeable of available statistical methods and be more careful with their selection of statistical tests.

However, I should also state that it is gratifying to see that the number of in-house quantitative research methods courses has recently increased (see comparison between Lazaraton et al., 1987, Loewen et al., 2014, and this study) but it is still not sufficient

compared to other sister disciplines such as education and psychology. Therefore, as Plonsky (2015) clearly stated, the field of SLA should provide more “*in-house* instruction on statistical techniques using sample data and examples tailored to the variables, interests, measures, and designs particular to L2 research” (p. 4).

In discussing the challenges that SLA doctoral students commonly face, most of the statistical conundrums also appeared to be pertinent to the inadequacy of application-based, field-specific statistical training. Put it simply, although students might be (knowledge-wise) whizzes in the implementation of a variety of statistical procedures, they might be clueless about what type of statistical tests would be more appropriate for their research questions. In support of this finding in the literature, Quilici and Mayer (1996), investigating the role of examples in how educational psychology students categorizing statistics problems, noted that:

Students in introductory statistics courses are expected to solve a variety of word problems that require using procedures such as *t* test, chi-square, or correlation. Although students may learn *how* to use these kinds of statistical procedures, a major challenge is to learn *when* to use them (p. 144).

On a related note, as one of the interviewees explicitly stated, SLA students, probably more statistically-naïve ones, might simply choose the statistical method they know best when they couldn't decide what tests to use. (This finding is actually in line with the results of the second research question in that students had slightly low performance on items asking participants to choose the statistical test appropriate for the scenario [e.g., S3Q12 and S5Q26] on the *SLA for SLA* survey). However, as Plonsky (2015) warned, “our analyses must be guided by substantive interests and relationships in question and not the other way around” (p. 4). It is important to have a broad statistical

repertoire but probably what is more important is to be able to know when to use them properly (Brown, 2015).

Regarding the proper use of statistics, this study showed that there were slightly different reporting practices among SLA doctoral students. Although it appeared to be common to follow the reporting standards of the publication manual of the American Psychological Association (APA, 2010), some students also reported drawing on other, probably easier, ways of reporting results of statistics as the primary basis such as following the reporting style of a published L2 study. In addition, some stated that they found it challenging to decide what to report and what to exclude in their paper, and thus some information that might be highly valuable, especially for meta-analysts, might go unreported. In fact, several L2 researchers in the field (Larson-Hall & Plonsky, 2015; Norris & Ortega, 2000; Norris et al., 2015; Plonsky, 2011, 2013; Plonsky & Gass, 2011) have raised concerns regarding inadequate reporting practices in the field. This issue might be a direct result of limited L2-specific guidelines particularly for employing new and more sophisticated statistical techniques. However, it is also worth noting that the field is slowly forming its own standards of reporting for quantitative research (see Larson-Hall & Plonsky, 2015). However, of course, reporting is also highly dependent on where publication takes place.

Statistical software packages (e.g., SPSS, *R*) make conducting many statistical techniques a lot easier. This study showed that SPSS was the most frequently used computation method, followed by *Excel* and *R*. However, how well SLA graduate students could use such tools is questionable. As pointed out in the interviews, the primary concern related to statistical software use is to rely heavily on default options

when performing statistical tests. Nevertheless, default options do not always produce the best results for certain tests (e.g., factor analysis, Plonsky & Gonulal, 2015). It is important to first have data screening (e.g., checking assumptions, detecting outliers) to have a sense of data, and then choose appropriate options.

Another interesting finding that I want to discuss is related to “third methodological movement” (Teddie & Taskakori, 2003, p. 5), following the developments of two somewhat opposite camps of research methodology (i.e., quantitative and qualitative). Although some L2 researchers are at one end of the continuum and some at the other, combining these two research methods in a single study, which is called mixed-methods research, is also becoming popular among L2 researchers (Gass, 2015). Indeed, in this seemingly highly quantitative study, I also made use of some elements of qualitative research method to have a better understanding of SLA doctoral student’s statistical literacy and statistical training as well. Considering the complex L2 phenomena that SLA doctoral students will probably deal with when they embark on their own research, it is important for them to be equipped with not only quantitative and qualitative research skills but also with mixed methods research skills because, as Leech and Huag (2015) emphasized, “graduating students from advanced education programs without an assurance of adequate research toolkit may be a disservice to them and to the field” (p. 105). Perhaps, it is time for SLA programs to require, or at least encourage, graduate students to take mixed methods research courses along with quantitative and qualitative research methods courses because taking quantitative or qualitative research methods courses will definitely help students develop skills in

dealing with quantitative or qualitative data, but these different courses may not be adequate enough to carry out mixed data analyses (Leech & Onwuegbuzie, 2010).

Before going on, it is necessary to reconsider the definition of statistical literacy in light of the results from this study, and to redefine it taking a somewhat broad approach. The findings in the present study highlighted that statistical literacy is more than the ability to read, understand and interpret statistical information presented in tabular and graphical format, but the ability to (a) choose correct statistical methods suitable for research questions, (b) conduct statistical analyses properly, (c) understand and interpret the results of statistical analyses, (d) evaluate the soundness of statistical analyses, and (e) report statistical results properly.

4.5 Limitations

In this study, I took an important step of examining statistical knowledge of doctoral students to provide a snapshot of the state of statistical literacy and training among SLA doctoral students, and where the field appears to be moving. However, the findings of this study should be interpreted with caution due to several limitations that might, to some extent, be attributed to the novel nature of the study. First and foremost, although the *SLA for SLA* survey focused on a variety of inferential statistics, relatively advanced and novel statistical tests (e.g., cluster analysis, Rasch analysis, Bayesian statistics) were not included to make the survey more manageable and to reach more SLA students. Future research would do well to use a more comprehensive survey covering not only descriptive statistics and common inferential statistics but also more advanced statistics, maybe using the *SLA for SLA* survey as a basis to better understand the statistical literacy among SLA researchers.

Also related to the design and content of future statistical literacy surveys is to include some *worry questions* (for a review, see Gal, 2002; for sample worry questions, see appendix G) about statistical results by using real L2 research examples. These types of questions can enable SLA researchers to ponder (a) how the data were collected, (b) the reliability of the instruments used, (c) how the data were analyzed, (d) what kinds of statistical tests were used and whether these tests were appropriate for the research questions, and (e) whether the results were interpreted properly. These types of questions would provide valuable information about SLA researchers' ability to critically question published L2 research. Further, future statistical literacy research might take statistics phobia or statistics anxiety into consideration when examining statistical knowledge as it can have debilitating effects on performance on statistical tests.

Finally, the sample was drawn from North America, and thus the findings might hold less import in other countries where the focus and amount of statistical training offered by SLA programs might be different.

4.6 Suggestions for the Field of SLA

In spite of several limitations discussed above, this study provided a state-of-the-art overview of current knowledge on statistics among SLA doctoral students, and highlighted some problems regarding statistical training in the field of SLA in North America. Rather than focusing solely on the problematic areas, I prefer to look ahead and to consider how the use of quantitative methods might be improved in the field. With the hope that this statistical literacy study and the issues raised here will encourage SLA graduate students, statisticians³, and SLA program directors to take more responsibility

for improving statistical literacy in the field, I offer some recommendations in the following section.

Before moving on to the recommendations, I should state that I am by no means arguing in this study that quantitative research methods are superior to qualitative research methods or all SLA doctoral students should be statisticians. In fact, I strongly agree with the recommendation of Wilkinson and APA Task Force on Statistical Inference (1999) regarding choosing “minimally sufficient” statistical analyses in research studies (p. 598). However, what I would like to emphasize here is that it is equally vital for SLA researchers to be aware of the available statistical procedures and techniques, and possess appropriate level of statistical knowledge in order to deal with the complex nature of questions posed in L2 research.

4.6.1 Improve statistical training in SLA

Improving statistical training in the field of SLA can be achieved through different ways. However, I believe that responsibility for improving statistical literacy in the field rests, to a great extent, on SLA programs because, although it may not be a ground-breaking discovery, this study showed that taking statistics courses is one of the efficient ways of improving statistical knowledge. Therefore, one simple suggestion for SLA programs (at least for larger, if not all, SLA programs) is to upgrade their curricular content to require more statistics courses. Of course, some SLA students may be inclined more to qualitative training than statistical training. Therefore, they may find it burdensome to take additional required statistics courses. Given that, offering such statistics courses as electives might be a valuable initial step. Also related to statistics courses, some of them appear to be taught in a more theoretical manner, with a limited

focus on hands-on or real-life data (e.g., choosing statistical methods appropriate for research questions). But it is necessary for statistics courses offered by SLA programs or outside departments to be able to prepare L2 researchers for the demands of real-life data (e.g., assumption violation, missing data, software issues).

Further, SLA programs may try to benefit from alumni feedback in regards to improve the quality of statistical training. Probably, recent graduates are “a highly credible group of program raters” (Morrison, Rudd, Zumeta & Nerad, 2011, p. 536), because they can provide some insightful suggestions regarding the quality of training in light of their experiences as students and newly-minted professors.

4.6.2 Increase the number of SLA faculty specializing in statistics

As explicitly stated by several interviewees, intermediate and advanced statistics courses are rarely offered by SLA programs, and thus students are usually sent to outside departments for higher-level statistical training. However, the content (i.e., examples and data sets used) of such outside courses is not always necessarily applicable to L2 research. It is therefore important to provide more in-house statistical training addressing the needs of L2 researchers.

Nonetheless, it is important to note here that there are not many SLA faculty who can teach such discipline-specific statistics courses. Considering the methodological and statistical reform movement taking place in applied linguistics (Plonsky, 2015), and introduction of novel and more sophisticated statistical methods to the field, this point becomes more important. Although these might be long-term goals, SLA programs may thus put more emphasis on training SLA professors, along with offering more courses for

SLA students. Further, it is important for those who regularly mentor doctoral students to have the necessary knowledge and skills themselves.

4.6.3 Increase students' awareness of quantitative methods for SLA

However, given the variety of statistical methods and the rise in the use of relatively advanced and novel statistical methods, it is not easy for SLA researchers to be highly knowledgeable in any statistical methods by just taking required statistics courses. It is possible, and probably easier now, to develop and improve statistical knowledge through self-training by making use of a variety of sources. For instance, there is a growing number of article- and book-length discipline-specific statistics sources (e.g., Larson-Hall, 2015; Plonsky, 2015; Loewen & Plonsky, 2015). In addition, several conferences in the field (e.g., American Association for Applied Linguistics [AAAL], Second Language Research Forum [SLRF]) have been offering statistics oriented workshops for SLA researchers (e.g., Statistics for applied linguistics with R' bootcamp led by Stefan Gries at SLRF in 2015). AAAL's recently-added research methods conference strand might be another way to see where the field is moving in terms of quantitative research methods.

Although I consider such efforts quite helpful and necessary, I am not optimistic about the number of students who are aware of and attend such workshops, seminars or conference strands. Therefore, I think more student-oriented environment focusing on methodological issues and developments is needed. In other words, students should be able to engage in research apprenticeship in quantitative L2 research. For instance, to my knowledge, two SLA programs have monthly statistics discussion meetings organized by graduate students with the support of quantitative-oriented faculty. In such meetings, the

use of relatively underused or sophisticated statistical methods are discussed. Probably, another important recommendation would be to encourage SLA graduate students to take more part in review process, at least in peer review, so that they can have some opportunities to hone their skills to critically question L2 quantitative research.

CHAPTER 5: CONCLUSION

This dissertation makes an important contribution to our understanding of the current state of statistical knowledge and statistical training among second language acquisition doctoral students, an area that we know so little about. In doing so, the present study highlighted problems pertinent to statistical training, and challenges in using statistical methods properly.

This study showed that although there is a slight increase in in-house statistical training in the field, the number of discipline-specific intermediate and advanced statistics courses is still limited. The current study also indicated that even though SLA doctoral students are good at understanding statistical information related to descriptive and inferential statistics, they find it challenging to interpret statistical results that are typically encountered in L2 research. The situation might be even worse when it comes more sophisticated and novel statistical methods. This is certainly an area worthy of the attention of future research.

Indeed, this study provides a strong basis for future studies into this important line of research. Given the important and continuing role that quantitative analysis plays in L2 research, and the complexity of L2 phenomena, it is critical for SLA researchers to be better equipped with necessary knowledge and skills to advance L2 theory and practice. Hopefully, the findings of this study would motivate graduate students, statisticians and SLA programs to take more concrete actions to move the field forward.

NOTES

NOTES

¹ Although there is a debate about SLA vs. applied linguistics, in this paper I just refer to the whole field as SLA, which in this paper encompasses SLA, applied linguistics, language assessment and testing.

²In this paper, SLA and L2 research are used interchangeably.

³I coined this term to describe SLA researchers who are highly knowledgeable in applied statistics and well-trained to use an array of statistical techniques properly within L2 research.

APPENDICES

APPENDIX A

SLA and Applied Linguistics Programs

Table 28

List of doctoral programs conferring degrees in SLA and applied linguistics

Institution	Department/Program Name
1. Arizona State University	Linguistics & Applied Linguistics
2. Carnegie Mellon University	Second Language Acquisition
3. Columbia University	Applied Linguistics & TESOL
4. Concordia University	Applied Linguistics
5. Georgetown University	Applied Linguistics
6. Georgia State University	Applied Linguistics & ESL
7. Indiana University- Bloomington	Second Language Studies
8. Iowa State University	Applied Linguistics & Technology
9. Northern Arizona University	Applied Linguistics
10. New York University- Steinhardt	TESOL
11. McGill University	Second Language Education
12. Michigan State University	Second Language Studies
13. Ohio State University	Foreign, Second and Multilingual Lang. Ed.
14. Penn State University	Applied Linguistics
15. Temple University	Education/Applied Linguistics
16. York University	Linguistics & Applied Linguistics
17. University of Alberta	Applied Linguistics
18. University of Arizona	Second Language Acquisition and Technology
19. University of British Columbia	Teaching English as a Second Language
20. University of Florida	Second Language Acquisition and Technology
21. University of Hawai'i	Second Language Studies
22. University of Illinois at Urbana-Champaign	Second Language Acquisition and Teacher Ed.
23. University of Iowa	Foreign Language & ESL Ed.
24. University of Maryland	Second Language Acquisition
25. University of Pennsylvania	Educational Linguistics
26. University of Pittsburgh	Linguistics with SLA orientation
27. University of Purdue	Second Language Studies
28. University of South Florida	Second Language Acquisition and Technology
29. University of Toronto	Applied Linguistics
30. University of Wisconsin	Second Language Acquisition

APPENDIX B

Background Questionnaire

1. Age _____

2. Gender: Male __ Female__

3a. What is your current academic position?

- MA student
- PhD student
- Other (Please specify)

3b. What year are you in your program? _____

3c. What is your major field of study?

- Applied Linguistics
- TESOL/TEFL
- Second Language Acquisition
- Foreign Languages
- Language Testing
- Education
- English
- Other _____

3d. What is your main research interest? _____

3e option1. What is the name of your current academic institution? _____

3e option 2. If you don't want to specify the name of your current academic institution, please click on the state where your institution is located.

- Psychology
- Linguistics
- Applied Linguistics
- Education
- Statistics
- Other _____

6a. Please rate the amount of training you have received in each category below.

Basic descriptive statistics (e.g., mean, median, standard deviation)

Very limited	Optimal
1 2 3 4 5	6

Common inferential statistics (e.g., t-test, ANOVA, chi-square, regression)

Very limited	Optimal
1 2 3 4 5	6

Advanced statistics (e.g., factor analysis, structural equation modeling, Rasch analysis, cluster analysis)

Very limited	Optimal
1 2 3 4 5	6

6b. To what extent are you satisfied with the amount of overall statistical training you have received?

Not satisfied at all	Very satisfied
1 2 3 4 5	6

7. To what extent do you do self-training in statistics/quantitative analysis?

Not at all	Exclusively
1 2 3 4 5	6

8. How frequently do you use the following sources to improve your statistical knowledge?

	Never				Very Frequently	
	1	2	3	4	5	6
Statistical textbooks	1	2	3	4	5	6
University Statistics Help Center	1	2	3	4	5	6
Statistics workshop	1	2	3	4	5	6
Professional consultants	1	2	3	4	5	6
Internet	1	2	3	4	5	6
Other colleagues	1	2	3	4	5	6
Other: _____	1	2	3	4	5	6

9. How do you compute your statistics? (Please select all that apply)

SPSS

R

SAS

Excel

STATA

AMOS

By hand

Other

I don't compute statistics

10. How statistically literate do you consider yourself?

Beginner

1

2

3

4

5

Expert

6

APPENDIX C

The SLA for SLA Instrument

The purpose of this survey is to examine the statistical knowledge of doctoral students in second language acquisition, applied linguistics or related programs in North America.

The survey consists of two main parts: a) a statistical background questionnaire and b) a statistical literacy assessment (SLA) survey. The SLA survey includes five scenarios that might be encountered in second language research, and twenty-eight multiple-choice questions related to these scenarios. The survey takes about 30 minutes to complete.

Even if you are not particularly quantitatively oriented, your responses will provide valuable information. All information will be stored confidentially, and you may discontinue the survey at anytime.

If you agree to take the survey, you will be compensated \$10 Amazon gift card for the survey. In addition, your results will be provided at the end of the survey. At the bottom of the results page at the end of the survey, you will see a link to receive your gift card. Please click on the link at the end of the survey and leave your email address to receive your gift card (Your email will not be linked to your survey responses). If you are also interested in participating in a follow-up interview, you will be compensated another \$10 Amazon gift card for the interview. Gift cards will be delivered via e-mail. Please don't use any additional sources when answering the questions.

If you have concerns or questions about this study, please contact the researcher (Talip Gonulal, Michigan State University, Second Language Studies Program, B-430 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, gonulalt@msu.edu, 614-440-1029) or the principal investigator (Dr. Shawn Loewen, Michigan State University, Department of Linguistics and Languages, B-255 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, loewens@msu.edu, 517-353-9790).

Thank you for your participation.

If you agree to take the survey, please select the 'Agree' option below and then click on the arrow.

- Agree
- Disagree

Scenario-1: Grammar instruction in English language classrooms

An English language center collected data from 2,581 English language learners (ELLs) at 50 different language institutions; institutions and ELLs were randomly selected to participate. To determine “what proportion of ELLs think that grammar instruction is necessary in English education,” ELLs were asked whether they thought grammar instruction was important. A total of 2,189 ELLs voted yes, and 392 ELLs voted no.

1. The sample is

- a. the 392 ELLs who voted no
- b. the 2,189 ELLs who voted yes
- c. the 2,581 ELLs in the study
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

2. The population is

- a. all ELLs in the world
- b. ELLs who think that grammar instruction is important
- c. ELLs who do NOT think that grammar instruction is important
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

3. Which of the following statements is TRUE?

- a. Descriptive statistics can provide information about the sample, and inferential statistics can provide information about the population.
- b. Descriptive statistics can provide information about the population, and inferential statistics can provide information about only the sample.
- c. Descriptive statistics can provide information about the parameter, and inferential statistics can provide information about the population.
- d. I don't know

Confidence: (Not confident) 1 2 3 4 5 6 7 8 9 10 (Confident)

Scenario-2: Language-related episodes in task-based activities

Part-I: A group of interactionist researchers investigate the number of language-related episodes (LREs) produced by 8 dyads during three different tasks (i.e., picture differences task, consensus task, and map task). The table below shows a subset of the raw data for the consensus task.

Table 29

The raw data for the consensus task

Dyad ID	1	2	3	4	5	6	7	8
Consensus task	0	5	2	17	3	2	1	2

4. The researchers calculate the mean, median and mode. One of the values they find is 2. What does the value 2 represent?

- a. The value of the mean, but not the median or mode
- b. The value of the median and the mode, but not the mean
- c. The value of the mean, median and mode
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

5. Based on this data set, which of the following options would be best to use to summarize the consensus task data?

- a. Use the most common number, which is 2
- b. Add up the 8 numbers in the bottom row and take the square root of the result
- c. Remove number 17, add up the other 7 numbers and divide by 7
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

6. If the standard deviation of the new consensus data is 1, which of the following statements would give the best interpretation of standard deviation?

- a. All of the LREs are one point apart
- b. The difference between the highest and the lowest number of LREs is 1 point
- c. The majority of LREs fall within one point of the mean
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Part-II: The table below shows the descriptive statistics for all three tasks.

Table 30

Descriptive statistics for all three tasks

	Mean	Median	Mode	SD	95% Confidence Intervals	
					Lower bound	Upper bound
Picture difference task	7.09	8	9	3.91	[5.03 - 9.15]	
Consensus task	4.00	2	2	1.00	[2.36 - 4.88]	
Map task	6.23	9	11	5.61	[6.17 - 10.29]	

7. Which of the following statements is TRUE?

- a. The variance in the map task data is the highest
- b. The variance in the picture difference task data is the highest
- c. The variances in the picture difference task data and the map task data are the same
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

8. Choose the graph that best represents the map task data.

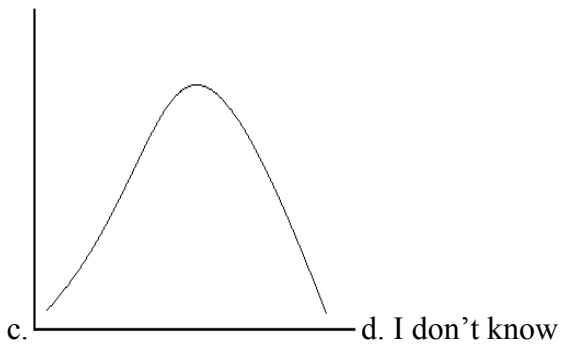
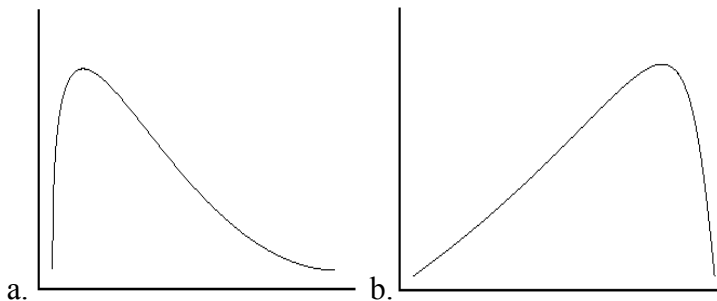


Figure 11. Graphs for map task data

Part III: Use the following boxplots to answer Questions 9-10

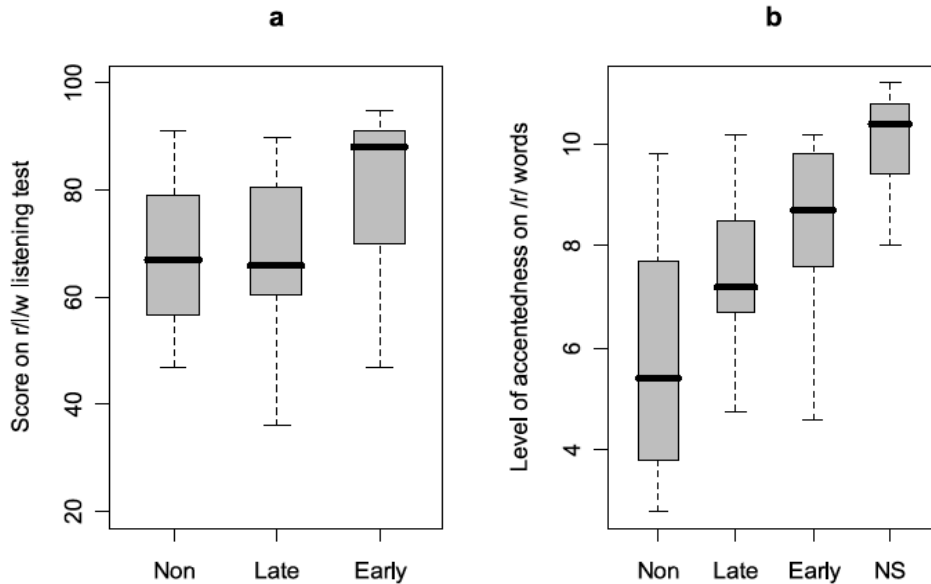


Figure 12 Boxplots for questions 9 and 10

9. Which is the best interpretation of the homogeneity of variance assumption based on these box-plots?

- a. Graph a shows similar variance among the three groups.
- b. Graph b shows similar variance among the four groups.
- c. Both graphs show similar variance among the groups.
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

10. What does the solid line in the middle of the box-plots represent?

- a. Mean
- b. Median
- c. Mode
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Scenario-3: Learners' choice of foreign language to study

Part -I: An English language program offers three unconventional foreign language courses (i.e., Dothraki, Klingon, and Esperanto). An L2 researcher working at this English language center is interested in studying whether male and female students differ in their choices of foreign language to study. The researcher counts how many male and female students are in each of these three courses. The researcher uses a statistical test to

investigate if there is a relationship between gender and the choice of foreign language to study.

11. Identify the type of variables in this study.

- a. Categorical
- b. Continuous
- c. Ratio
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

12. Choose the statistical test that is the most appropriate for this research study.

- a. Paired sample t-test
- b. Repeated measures analysis of variance
- c. Chi-square
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Part-II: After data screening and testing the assumptions, the researchers decide to use a chi-square test to investigate if there is a relationship between gender and the choice of foreign language to study (i.e., Dothraki, Klingon, and Esperanto). The results of the chi-square test are $X^2(2, n=50) = 2.10, p = .58, \text{Cramer's } V = .09$ (alpha level set at .05).

13. Which of the following statements is TRUE?

- a. There is no statistical relationship between gender and the choice of foreign language to study
- b. There is a statistical relationship between gender and the choice of foreign language to study
- c. The choice of foreign language studied can be statistically determined by gender
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

14. If the probability of making a type II error in this study is 0.15, what is the power of the analysis?

- a. .85
- b. 1.15
- c. The power cannot be determined based on this information
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

15. If the sample size of the study was 100 instead of 50, how would the power of the study be affected?

- a. It would increase

- b. It would decrease
- c. It would not be affected
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

16. Which of the following statements is TRUE about the effect size of this study?

- a. It has a small effect size
- b. It has a medium effect size
- c. It has a large effect size
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Scenario-4: Vocabulary learning in a second language

Part-I: A group of L2 researchers investigate whether the amount of formal instruction (in weeks) that a bilingual student receives matters to how many words they will learn in Spanish. They conduct a statistical test to examine the possible relationship between the amount of formal instruction and amount of vocabulary learned in Spanish.

17. Identify the type of variables in this study

- a. Categorical
- b. Continuous
- c. Dichotomous
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

18. Choose the statistical test that is the most appropriate for this research study

- a. Paired sample t-test
- b. Correlation
- c. Factor analysis
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Part-II: The researchers conduct a correlation test to examine the possible relationship between the amount of formal instruction ($M = 22.7$, $SD = 4.3$) and amount of vocabulary learned in Spanish ($M = 45.4$, $SD = 8.1$). The results of the correlation are $n = 66$, $r = .89$, 95% CI [.82, .93], $r^2 = .79$, $p = .04$.

19. Which of the following statements is TRUE?

- a. The relationship between two variables is statistically significant, positive and strong

- b. The relationship between two variables is statistically significant and positive but weak
- c. The relationship between two variables is positive and strong but not statistically significant
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Label each type of statistic:

20. $M = 22.7$ a. Descriptive b. Inferential c. Both d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

21. $SD = 8.1$ a. Descriptive b. Inferential c. Both d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

22. $r = .89$ a. Descriptive b. Inferential c. Both d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

23. $p = .04$ a. Descriptive b. Inferential c. Both d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

24. What type of error would the researchers have committed if the statistically significant correlation they found was actually a false positive?

- a. Type I error
- b. Type II error
- c. Standard error
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

25. If the statistical coefficient in this study has a high standard error, which of the following statements would be TRUE?

- a. The difference between the population correlation coefficient and the sample correlation coefficient is large
- b. The difference between the population correlation coefficient and the parameter correlation coefficient is small
- c. The difference between the population correlation coefficient and the parameter correlation coefficient is large
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Scenario-5: Factors affecting tonal accuracy in a second language

Part-I: An L2 researcher is interested in studying how individual factors (i.e., language aptitude, age, motivation level, type of instruction, and amount of instruction) result in higher levels of tonal accuracy in second language learners of Thai. The researcher

examines how much of the differences in scores on a tone test can be explained by these five items.

26. Choose the statistical test that is the most appropriate for this research study

- a. Multiple regression
- b. Factor analysis
- c. Kruskal Wallis
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

Part-II: The table below shows the relationship between the level of tonal accuracy in Thai and the five predictor variables (i.e., language aptitude, age, motivation level, type of instruction, and amount of instruction) for the three groups of participants.

Table 31

The results of the multiple regression analysis

	N	R	R ²	F	Sig.
Advanced learners	30	.96	.92	67.00	.00
Intermediate learners	30	.75	.56	84.31	.06
Beginner learners	30	.65	.42	91.49	.20

27. Which of the following statements is TRUE?

- a. There is a statistically significant relationship between the level of tonal accuracy and the five predictor variables for the intermediate learners.
- b. There is a statistically significant relationship between the level of tonal accuracy and the five predictor variables for the advanced learners
- c. There is a statistically significant relationship between the level of tonal accuracy and the five predictor variables for the beginner learners
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

28. Which of the following statements is TRUE?

- a. The five predictor variables explain 56% of the variation in the level of tonal accuracy among the intermediate learners
- b. The five predictor variables explain 67% of the variation in the level of tonal accuracy among the advanced learners
- c. The five predictor variables explain 20% of the variation in the level of tonal accuracy among the beginner learners
- d. I don't know

Confidence: (Not confident at all) 1 2 3 4 5 6 7 8 9 10 (Very confident)

1. Did you use any additional source when answering the questions on this survey?

Yes__ No__

If yes, which of the following sources did you use for statistical assistance?

Statistical textbook

Other colleagues

Internet

Other _____

Calculator

2. Could you please give me your impressions of the survey you completed? How well do you think you did on the survey?

3. Is there anything that you would like to tell me about your experience with statistical analyses and your training in statistics/quantitative research methods?

Thank you for taking the survey!

APPENDIX D

Interview Questions

Performance on the SLA Survey

1. How well do you think you did in the statistics test?
2. Which questions / scenarios did you find easy? Why?
3. Which questions / scenarios gave you the most difficulty? Why?
4. How relevant do you think the questions/scenarios are to your research experience and statistical training?

Statistical Training

5. Could you describe your personal development in terms of quantitative research methods within SLA research?
6. Could you tell me about the different types of training you have received on how to perform statistical analyses?
 - What is the total number of quantitative research methods/statistics courses required in your program?
 - How many quantitative research methods/statistics courses have you taken? Which department(s) offered those courses?
 - What resources does your university provide for you to maintain your statistical knowledge? Do you take advantage of these opportunities?
 - Are there any statistical concepts and procedures that you wish to receive further training?
7. How informed do you feel you are about best practices in statistical analyses?

Experiences with Statistics

8. How often do you incorporate statistical procedures and concepts in your research?
9. Could you share some of the difficulties you have faced while performing statistical analyses?
 - What resources do you rely on for assistance when facing difficulties (e.g., when you are unsure of what statistical method you need to use, what and how to report)?
10. Could you share a little about your most recent statistical conundrum?
11. How often do you read the analysis and results sections of papers, as opposed to going straight to the discussion section? Do you sometimes disagree with the type of analysis researchers performed or with the conclusions they drew based on their findings?
12. What is your overall impression of the statistical knowledge of SLA graduate students in general?

APPENDIX E

Survey Invitation Email

Dear Professor X,

I am a PhD candidate in Second Language Studies program at Michigan State University. As part of my dissertation, I am conducting a study on the statistical knowledge and training of doctoral students in second language acquisition, applied linguistics or related programs in North America.

I am currently recruiting participants for my study and was hoping if you could distribute the following survey invitation to doctoral students in your program.

Thank you very much for your time.

Best,

Talip

=====

Dear Doctoral Student,

My name is Talip Gonulal. I am a PhD candidate at Michigan State University. As part of my dissertation research, I am examining the current state of statistical knowledge of doctoral students in second language acquisition, applied linguistics or related programs in North America. In addition, I am interested in what training graduate students in the field have received in quantitative research methods.

I would like to invite you to participate in this study by completing an online survey. The survey consists of two main parts: a) a statistical background questionnaire and b) a statistical literacy assessment (SLA) survey. The SLA survey includes five scenarios that might be encountered in second language research, and twenty-eight multiple-choice questions related to these scenarios. The survey takes about 30 minutes to complete. All information will be stored confidentially, and you may discontinue the survey at anytime. Your participation is highly appreciated even if you are not particularly quantitatively oriented.

If you agree to take the survey, you will be compensated \$10 Amazon gift card for the

survey. In addition, your results will be provided at the end of the survey. Please click on the link at the end of the survey and leave your email address to receive your gift card (Your email will not be linked to your survey responses). If you are also interested in participating in a follow-up interview, you will be compensated another \$10 Amazon gift card for the interview. Gift cards will be delivered via e-mail.

If you have concerns or questions about this study, please contact the researcher (Talip Gonulal, Michigan State University, Second Language Studies Program, B-430 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, gonulalt@msu.edu, 614-440-1029) or the principal investigator (Dr. Shawn Loewen, Michigan State University, Department of Linguistics and Languages, B-255 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, loewens@msu.edu, 517-353-9790).

By clicking on the following link, you agree to take part in this survey:

https://broad.qualtrics.com/SE/?SID=SV_0BSF23PAQp3Yloh

I would be grateful if you could forward this email to whoever you think may be interested.

Thank you in advance for your time!

Sincerely,

Talip Gonulal

Apologies for cross-posting

=====

APPENDIX F

Interview Invitation Email

Dear Researcher,

Thank you for taking the statistical literacy survey. In the survey, you expressed your interest in participating in a follow-up interview.

I am now setting up interviews for the follow-up and would like to schedule an interview with you.

The interview takes 20-30 minutes and will be conducted via Skype. You will be compensated \$10 Amazon gift card for your time. I am simply trying to capture your experiences and training in quantitative research methods. The information you provide will be completely confidential and used for research purposes only.

Please let me know what day and time works best for you and I'll do my best to be available. If you have any questions, please do not hesitate to ask.

I look forward to hearing from you.

Best,

Talip Gonulal

APPENDIX G

Sample Worry Questions about Statistical Messages (Gal, 2002)

1. Where did the data (on which this statement is based) come from? What kind of study was it? Is this kind of study reasonable in this context?
2. Was a sample used? How was it sampled? How many people did actually participate? Is the sample large enough? Did the sample include people/units which are representative of the population? Is the sample biased in some way? Overall, could this sample reasonably lead to valid inferences about the target population?
3. How reliable or accurate were the instruments or measures (tests, questionnaires, interviews) used to generate the reported data?
4. What is the shape of the underlying distribution of raw data (on which this summary statistic is based)? Does it matter how it is shaped?
5. Are the reported statistics appropriate for this kind of data, e.g., was an average used to summarize ordinal data; is a mode a reasonable summary? Could outliers cause a summary statistic to misrepresent the true picture?
6. Is a given graph drawn appropriately, or does it distort trends in the data?
7. How was this probabilistic statement derived? Are there enough credible data to justify the estimate of likelihood given?
8. Overall, are the claims made here sensible and supported by the data? e.g., is correlation confused with causation, or a small difference made to loom large?

9. Should additional information or procedures be made available to enable me to evaluate the sensibility of these arguments? Is something missing? e.g., did the writer "conveniently forget" to specify the base of a reported percent-of-change, or the actual sample size?

10. Are there alternative interpretations for the meaning of the findings or different explanations for what caused them, e.g., an intervening or a moderator variable affected the results? Are there additional or different implications that are not mentioned?

REFERENCES

REFERENCES

- Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist, 63*(1), 32-50.
- Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. *American Psychologist, 45*(6), 721-734.
- Allen, K. (2006). The statistics concept inventory: Development and analysis of a cognitive assessment instrument in statistics. (Unpublished doctoral dissertation). University of Oklahoma, Norman, OK.
- Bailey, K. M. & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236–256). Clevedon, UK: Multilingual Matters.
- Becker, B. J. (1996). A look at the literature (and other resources) on teaching statistics. *Journal of Educational and Behavioral Statistics, 21*(1), 71-90.
- Ben-Zvi, D. & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. In D. Ben-Zvi & J. B. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishing.
- Borders, L. D., Wester, K. L., Fickling, M. J., & Adamson, N. A. (2014). Research training in doctoral programs accredited by the council for accreditation of counseling and related educational programs. *Counselor Education and Supervision, 53*(2), 145-160.
- Brown, J. D. (2004). Resources on quantitative/statistical research for applied linguists. *Second Language Research, 20*(4), 372-393.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw-Hill.
- Brown, J. D. (2013). Teaching statistics in language testing courses. *Language Assessment Quarterly, 10*(3), 351-369.

- Brown, J. D. (2015). Why bother learning advanced quantitative methods in L2 research? In L. Plonsky (Ed), *Advancing quantitative methods in second language research*. New York: Routledge.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349-383.
- Capraro, R. M., & Thompson, B. (2008). The educational researcher defined: What will future researchers be trained to do? *The Journal of Educational Research*, 101(4), 247-253.
- Chaudron, C. (2001). Progress in language classroom research: Evidence from The Modern Language Journal, 1916–2000. *Modern Language Journal*, 85, 57–76.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Los Angeles, CA: SAGE.
- Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. Los Angeles, CA: SAGE.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369-382.
- Curtis, D. A., & Harwell, M. (1998). Training doctoral students in educational statistics in the United States: A national survey. *Journal of Statistics Education*, 6(1).
- Dauzat, S. V., & Dauzat, J. (1977). Literacy: In quest of a definition. *Convergence*, 10(1), 37-41.
- Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge: Cambridge University.
- Estrada, A., Batanero, C., & Lancaster, S. (2011). Teachers' attitudes towards statistics. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Teaching statistics in school mathematics - Challenges for teaching and teacher education* (pp. 163-174). The Netherlands: Springer.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE.

- Finney, S., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*, 161–186.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25.
- Gal, I. (2004). Statistical literacy, meanings, components, responsibilities. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishing.
- Galesic, M., & Garcia-Retamero, R. (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine, 170*(5), 462-468.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*(1), 22-38.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372-396.
- Gass, S. (2009). A survey of SLA research. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 3–28). Bingley, UK: Emerald.
- Gass, S. (2015). Methodologies of second language acquisition. In M. Bigelow & J. Ennser-Kananen (Eds.), *The Routledge handbook of educational linguistics* (pp. 9–22). New York/London: Routledge/Taylor & Francis.
- Gass, S., Fleck, C., Leder, N., & Svetics, I. (1998). Ahistoricity revisited. *Studies in Second Language Acquisition, 20*(03), 407-421.
- Godfroid, A., & Spino, L. (2015). Reconceptualizing reactivity of think-alouds and eye-tracking: Absence of evidence is not evidence of absence. *Language Learning, 65*(4), 896-928.
- Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychology/Psychologie canadienne, 50*(2), 83.
- Gonulal, T., Loewen, S., & Plonsky, L. (in preparation). The development of statistical knowledge in second language research.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

- Gries, S. (2010). Methodological skills in corpus linguistics: A polemic and some pointers towards quantitative methods. In T. Harris & M. M. Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 121–146). Frankfurt, Germany: Peter Lang.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205.
- Henson, R. K., Hull, D. M., & Williams, C. S. (2010). Methodology in our education research culture toward a stronger collective quantitative proficiency. *Educational Researcher*, 39(3), 229-240.
- Henson, K. R., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66(3), 393-416.
- Huberty, C. J., Dresden, J., & Bak, B. G. (1993). Relations among dimensions of statistical knowledge. *Educational and Psychological Measurement*, 53(2), 523-532.
- Jeon, E. H. (2015). Multiple regression. In L. Plonsky (Ed), *Advancing quantitative methods in second language research*. New York: Routledge.
- Jones, F. R. (1998). Self-instruction and success: A learner-profile study. *Applied Linguistics*, 19(3), 378-406.
- Jones, M. (2013). Issues in doctoral studies – Forty years of journal discussion: Where have we been and where are we going? *International Journal of Doctoral Studies*, 8, 83-104.
- Kline, P. (1999). *The handbook of psychological testing*. London: Routledge.
- Kirsch, I., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Larson–Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R*. Routledge.
- Larson-Hall, J., & Herrington, R. (2010). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics*, 31(3), 368-390.

- Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(S1), 127-159.
- Lazaraton, A. (2000). Current trends in research methodology and statistics in applied linguistics. *TESOL Quarterly*, 34, 175-181.
- Lazaraton, A. (2005). Quantitative research methods. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 109–224). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lazaraton, A., Riggenbach, H., & Ediger, A. (1987). Forming a discipline: Applied linguists' literacy in research methodology and statistics. *TESOL Quarterly*, 21, 263–277.
- Leech, N. L., & Goodwin, L. D. (2008). Building a methodological foundation: Doctoral-level methods courses in colleges of education. *Research in the Schools*, 15(1), 1-8.
- Leech, N., & Haug, C. A. (2015). Investigating graduate level research and statistics courses in schools of education. *International Journal of Doctoral Studies*, 10, 93-111.
- Leech, N. L., & Onwuegbuzie, A. J. (2010). Epilogue: The journey: From where we started to where we hope to go. *International Journal of Multiple Research Approaches*, 4(1), 73-88.
- Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65(S1), 185-207.
- Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. New Jersey: John Wiley & Sons.
- Loewen, S., & Gass, S. (2009). Research timeline: The use of statistics in L2 acquisition research. *Language Teaching*, 42(2), 181-196.
- Loewen, S., & Gonulal, T. (2015). Principal component analysis and factor analysis. In L. Plonsky (Ed), *Advancing quantitative methods in second language research*. New York: Routledge.
- Loewen, S., & Plonsky, L. (2015). *An A-Z of applied linguistics research methods*. New York: Palgrave.
- 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(2), 360-388.

- Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design*. New York: Routledge.
- Morrison, E., Rudd, E., Zumeta, W., & Nerad, M. (2011). What matters for excellence in PhD programs? Latent constructs of doctoral program quality used by early career social scientists. *The Journal of Higher Education*, 82(5), 535-563.
- Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning*, 65(S1), 97-126.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417-528.
- Norris, J. M., Ross, S. J., & Schoonen, R. (2015). Improving second language quantitative research. *Language Learning*, 65(S1), 1-8.
- Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educational and Psychological Measurement*, 63(6), 1020-1038.
- Osborne, J. W. (2012). *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Los Angeles: Sage.
- Patil, V. H., Singh, S. N., Mishra, S., & Donovan, D. T. (2007). Parallel analysis engine to aid determining number of factors to retain. Computer software. Retrieved from <http://ires.ku.edu/~smishra/parallelengine.htm>.
- Pierce, R., & Chick, H. (2013). Workplace statistical literacy for teachers: Interpreting box plots. *Mathematics Education Research Journal*, 25(2), 189-205.
- 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, 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. (Ed.) (2015). *Advancing quantitative methods in second language research*. New York: Routledge.
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, 61(2), 325-366.

- 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.
- Plonsky, L., Egbert, J., & LaFlair, G. T. (2014). Bootstrapping in applied linguistics: Assessing its potential using shared data. *Applied Linguistics, 1*-21.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.
- Polio, C., & Gass, S. (1997). Replication and reporting. *Studies in Second Language Acquisition, 19*(4), 499-508.
- Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*(1), 144.
- Rossen, E., & Oakland, T. (2008). Graduate preparation in research methods: The current status of APA-accredited professional programs in psychology. *Training and Education in Professional Psychology, 2*(1), 42.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*(2), 147-177.
- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences, 3*, 153-160.
- Schield, M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance, 1*(1), 15-20.
- Schield, M. (2002). *Statistical Literacy Survey*. Retrieved from www.StatLit.org/pdf/2006SchieldIASSIST.pdf
- Schield, M. (2004). Statistical literacy and liberal education at Augsburg College. *Peer Review, 6*, 16-18. Retrieved from www.StatLit.org/pdf/2004SchieldAACU.pdf.
- Schield, M. (2006). Statistical literacy survey results: Reading graphs and tables of rates and percentages. *Conference of the International Association for Social Science Information Service and Technology (IASSIST)*.
- Schield, M. (2010). Assessing statistical literacy: Take CARE. In P. Bidgood, N. Hunt & F. Jolliffe (eds) *Assessment Methods in Statistical Education: An International Perspective*. John Wiley & Sons Ltd.
- Selinker, L., & Lakshmanan, U. (2001). How do we know what we know? Why do we believe what we believe? *Second Language Research, 17*, 323-325.

- Skidmore, S. T., & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement, 70*(5), 777-795.
- Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671-701). Thousand Oaks, CA: SAGE.
- The American Heritage Dictionary of the English language*. 4th ed. Boston, MA: Houghton Mifflin; 2000.
- Thomas, M. (2013). The doctorate in second language acquisition: An institutional history. *Linguistic Approaches to Bilingualism, 3*(4), 509-531.
- Thompson, B. (1999). Five methodology errors in educational research: A pantheon of statistical significance and other faux pas. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 23-86). Stamford, CT: JAI Press.
- Thompson, A., Li, S., White, B., Loewen, S., & Gass, S. (2012). Preparing the future professoriate in second language acquisition. *Working Theories for Teaching Assistant Development, 137-167*.
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1-8.
- Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education*. Amsterdam: IOS Press.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3-46.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-248.
- Winke, P. (2014). Testing hypotheses about language learning using structural equation modeling. *Annual Review of Applied Linguistics, 34*, 102-122.
- Yilmaz, M. R. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education, 4*(1), 1-9.
- Zimiles, H. (2009). Ramifications of increased training in quantitative methodology. *American Psychologist, 64*, 51-56.