Available online at www.ejal.eu

Eurasian Journal of Applied Linguistics 4(1) (2018) 49–70

EJAL
Eurasian Journal of Applied Linguistics

An Investigation of the Predictors of Statistical Literacy in Second Language Acquisition

Talip Gonulal [a] *

[a] *Erzincan University, English Language Teaching, Erzincan 24100, Turkey*

Received 18 January 2018 | Received in revised form 27 February 2018 | Accepted 26 March 2018

Abstract

The use of statistics in second language acquisition (SLA) research has increased over the past 30-40 years and continues to increase in both complexity and sophistication (Gass, 2009; Loewen & Gass, 2009). The increased use of statistical procedures has drawn attention to the current state of statistical literacy among second language (L2) researchers. Statistical literacy is a critical skill to acquire on the parts of both the producers and consumers of L2 research. However, it is a relatively new research topic in the field. So, little is known regarding what factors play key roles in the development of statistical literacy. Therefore, this study attempted to investigate the predictors of statistical literacy in SLA. One hundred and twenty SLA doctoral students took a statistical background questionnaire and a discipline-specific statistics survey. A series of multiple regression analyses were conducted on the statistics survey data. The results indicated that number of statistics courses taken, quantitative research orientation, and self-training in statistics were the significant predictors of statistical literacy. In light of the findings of this study, several suggestions directed toward improving statistical literacy in the field of SLA were made.

Keywords: Statistical literacy; statistical training; quantitative research orientation; SLA

1. Introduction

Although the field of second language acquisition (SLA) is relatively new, the use of quantitative research methods has been prevailing since its inception. However, the field has seen an exponential increase in the use of statistical procedures in the last two decades. To illustrate, 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 (L2) research has noticeably increased (Loewen & Gonulal, 2015; Plonsky, 2013, 2014, 2015; Plonsky & Gonulal, 2015; Winke, 2014). Along with the current trend towards the use of novel and more sophisticated statistical methods in L2 research, there is a growing number of article- and book-length sources (e.g., Larson-Hall, 2015; Mackey & Gass, 2015; Plonsky, 2015; Roever & Phakiti, 2018) dealing with discipline-specific statistics and

* Talip Gonulal.
E-mail address: talip.gonulal@erzincan.edu.tr
http://dx.doi.org/.....

quantitative research designs. Plonsky (2015) described this increasing trend as "methodological and statistical reform movement" (p. 4).

It is, however, important to mention here that 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, Ross & Schoonen, 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. More specifically, a statistically literate L2 researcher should be able 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, 2017).

1.1. Statistical literacy

Statistical literacy is a new research area in L2 research, although it has been investigated in other fields, mostly in statistics and mathematics education. 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 of 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).

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, statistical literacy within the domain of SLA was operationalized as 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.

Given the apparent significance of statistical literacy as a necessary skill to be acquired by SLA researchers, a few scholars (e.g., Gonulal et al., 2017; Lazaraton, Riggenbach & Ediger, 1987; Loewen et al., 2014) conducted studies on the state of statistical literacy among L2 researchers, including both professors and graduate students. However, little is still known regarding what kinds of factors are important in the development of statistical literacy. Unearthing such factors will definitely help L2 researchers improve their statistical literacy to "better address questions currently posed and to take on novel or more complex questions (Plonsky, 2015, p. 4). Therefore, following research question guided this study:

RQ 1: What factors predict SLA doctoral students' statistical literacy?

2. Method

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, the scope of the study was limited to North America. Of the approximately 900 graduate students contacted, 120 took part in this study. The participants were from approximately thirty universities across North America. 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.

2.2. Instruments

Data for this study came from two sources: (a) a statistical background questionnaire and (b) a statistical literacy assessment survey.

2.2.1. Statistical background questionnaire

In order to elicit information about participants' statistical training, a questionnaire was developed, 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 A).

2.2.2. Statistical literacy assessment for second language acquisition survey

In this study, a statistical literacy assessment for second language acquisition (henceforth, SLA for SLA) instrument was used. The SLA for SLA survey was originally designed by Loewen et al. (under review) to measure SLA researchers' knowledge of statistics. The SLA for SLA consisted of five scenarios and twenty-eight multiple-choice statistics questions (see Appendix B). After pilot testing the survey, Loewen et al. conducted an in-depth item analysis to examine the quality of the items on the survey. The results of the analyses showed that the SLA for SLA survey measures a single construct (i.e., statistical literacy). Further, the overall reliability of the survey (Cronbach's $\alpha = .92$) was quite high (Field, 2013; Kline, 1999).

2.3. Procedure

An online version of the *SLA for SLA* survey was created via *Qualtrics*. Then, a complete list of institutions offering doctoral degrees in SLA in North America were created, following Thompson, White, Loewen and Gass's (2012) list. Afterward, a survey invitation email was drafted and forwarded to several program directors and statistics instructors to share the link with doctoral students in their program. Personal invitation emails were also sent to doctoral students whose email addresses were listed on their programs' websites. To increase the rate of participation, participants were compensated with $10 gift cards. Overall, the data collection process took 13 weeks.

2.4. Data analysis

A series of multiple regression analyses were run. To get reliable multiple regression analysis results, the data were screened and the assumptions were checked (see Table 1). First, the sample size was examined to see if the data were appropriate for regression. According to Field (2013), there should be at least 15 participants for each predictor variable. Given that, the sample of 120 would be adequately large for a regression analysis with four predictor variables. Then, further data screening was conducted to see whether there were any univariate and multivariate outliers. To this end, the Mahalanobis distance which is fundamentally the distance of an item from the multivariate mean was computed (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.

Table 1. 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 (2013), and Tabachnick and Fidell (2013).

Further, the assumption of multicollinearity which can pose a real problem for multiple regression analysis was checked. Thus, the variance inflation factors (VIF) and tolerance values were examined 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, linearity and homoscedasticity (i.e., assumption of equal variance) were also checked by examining the scatter plots of variables and the residual plots. Overall, the results showed that the data were appropriate for multiple regression analyses.

3. Results

SLA doctoral students 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 Table 2. Regarding the research orientation, there were no significant differences in their preference over being quantitatively- or qualitatively-oriented. In other words, they reported conducting qualitative research as frequently as quantitative research. In response to the question regarding whether they do self-training in statistics, SLA doctoral students did not draw a positive picture (*M* = 3.00, *SD* = 1.41, *95% CI* [2.74, 3.26]). As for the SLA doctoral students' performance on the SLA for SLA survey results, the average overall score was 16.38 (*SD* = 7.82, *95% CI* [14.96, 17.79]) out of 28. This means that the survey was slightly challenging for most of the SLA doctoral students.

In order to find a good model that can predict SLA graduate students' statistical literacy, multiple regression analyses were run. For this purpose, hierarchical (sequential) regression was chosen using 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 this study investigates how different predictor variables would explain the variance in statistical literacy, while controlling for previously entered variables.

Table 2. Statistical background information

	N	M	SD	95% CI
Number of stats courses taken	118	2.19	1.56	[1.91, 2.48]
Research orientation				
Quantitative research orientation[a]	116	3.44	1.44	[3.17, 3.70]
Qualitative research orientation[a]	118	3.24	1.36	[2.99, 3.48]
Self-training in statistics[a]	117	3.00	1.41	[2.74, 3.26]

Note. [a]1 = Not at all, 6 = Exclusively

In hierarchical regression, the order of entry is often determined by theoretical or empirical importance (Field, 2013; Jeon, 2015). However, because this area of research has been relatively untapped in the field, the order of the predictor variables entered in the analyses were determined 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, self-training in statistics and years spent in a program was entered first, followed by other two variables. Tables 3 and 4 present the results of this analysis.

Table 3. Regression model summary for overall score

Model	*R*	R^2	Adjusted R^2	SEE	*F* change	*df1*	*df2*	Sig. *F* 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. [a]Number of courses; [b]Quantitative orientation; [c]Self-training; [d]Year in program.

The model with all four predictors accounted for 29.3% of the variance. Number of courses and quantitative orientation had significant positive regression weights, indicating participants with higher score on these variables were expected to perform better on the SLA for SLA survey. Indeed, these two were the best predictor variables, 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 to the model at all.

Table 4. Model data for overall score

		Std.				95%CI	
Model	B	error	*β*	*t*	Sig.	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

As for the alternative regression model where self-training and year in program went first, three out of four variables significantly contributed to the alternative model (see Tables 5 and 6). 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 year in program.

Table 5. Alternative regression model summary for overall score

Model	*R*	R^2	Adjusted R^2	SEE	*F* change	*df1*	*df2*	Sig. *F* 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. [a]Self-training; [b]Year in program; [c]Number of courses; [d]Quantitative orientation.

Table 6. Alternative model data for overall score

Model	B	Std. error	β	t	Sig.	95%CI Lower	95%CI 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.

4. Discussion and Conclusions

The primary purpose of this study was to investigate what factors would be predictive of statistical literacy. Presumably, many L2 researchers would easily suggest that number of courses taken in statistics alone is predictive of statistical literacy. Although a few studies (e.g., Gonulal et al., 2017; 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 statistical literacy. These findings suggest that as might be expected, SLA doctoral 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 relevant studies, similar findings have been reported. For instance, 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 pattern 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. (2017) 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, Hull & Williams, 2010) and psychology (Aiken, West, & Millsap, 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. 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. However, it is surprising that 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 papers, 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 an SLA program might be subsumed by courses and/or orientation. However, all these are speculative. Thus, further research is certainly needed in this area.

An important point to consider is that although the multiple regression analyses produced significant results regarding the predictors of statistical literacy, the total variance explained by the suggested model in this study was not very high (29.3%). One potential predictor that was not examined in this is study but could contribute to the model is the level of statistics anxiety. It is possible that statistics anxiety might have had debilitating effects on some participants' performance on the SLA for SLA survey. Indeed, in his

study using path analysis to develop a model to explain statistics achievement among graduate students in social and behavioral sciences, Onwuegbuzie (2003) reported that statistical achievement was negatively correlated with statistics anxiety.

In light of the results of the multiple regression analyses, it is important to note here that the field has made some progress in regards to the number of statistics courses taken over 2.5 decades (see Gonulal et al., 2017; Lazaraton et al., 1987; Loewen et al., 2014). Although this study indicated that taking statistics courses is an important way of developing statistical literacy, it does not necessarily ensure higher level statistical knowledge. The content of the statistical training is also equally important. Not surprisingly, the overall statistical training in the field seems to be limited to largely introductory, and partially intermediate concepts and procedures (Gonulal, under review). In the same vein, even though self-training is another important predictor of statistical literacy, self-training in statistics was not very common among SLA graduate students. Considering the factors playing key role in statistical literacy development and the current state of statistical training provided in the field, several informed suggestions can be made. One simple suggestion for SLA programs (at least for larger, if not all, SLA programs) would be to upgrade their curricular content to offer more statistics courses. Further, although these might be long-term goals, SLA programs may put more emphasis on statistical training of SLA professors because it is important for those who regularly mentor doctoral students to have the necessary knowledge and skills themselves. Finally, SLA students' awareness of quantitative research methods might be increased through conferences, workshops, peer review and research apprenticeship as well.

The present study provides important evidence regarding the factors that influence the development of statistical literacy in SLA. 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 met the necessary psychometric requirements to reliably measure the construct of statistical literacy, the focus of the survey was largely on basic statistics and common 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. In fact, this slightly narrow scope of the survey might also explain the low variance explained by the regression model. 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. Further, future statistical literacy research might take statistics phobia or statistics anxiety into consideration when examining statistical knowledge as it can have negative impact on statistical test performance. Finally, the data for this study came only 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.

In spite of these limitations, the current study provides useful information regarding the significant predictors of statistical literacy which is a necessary skill not only for producers but also consumers of L2 research. Indeed, considering the essential role of quantitative research methods in L2 research, it is important for L2 researchers in general and newly-minted PhDs in particular to be better equipped with necessary knowledge and skills to advance L2 theory and practice. Hopefully, the findings of this study would shed light on how L2 researchers can improve their statistical knowledge to move the field forward.

Acknowledgements

This study is part of my dissertation conducted at Michigan State University. Special thanks go to my dissertation committee members: Shawn Loewen, Susan Gass, Paula Winke and Aline Godfroid. I am also grateful to Michigan State University for funding this study. Many thanks also go to anonymous EJAL reviewers for their comments and suggestions.

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.

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

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.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. Los Angeles, CA: SAGE.

Dauzat, S. V., & Dauzat, J. (1977). Literacy: In quest of a definition. *Convergence*, *10*(1), 37-41.

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.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: SAGE.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25.

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.

Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychology/Psychologie canadienne*, *50*(2), 83-90.

Gonulal, T. (under review). Investigating the statistical knowledge and training of doctoral students in second language acquisition. *The Modern Language Journal*.

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

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.

Jeon, E. H. (2015). Multiple regression. In L. Plonsky (Ed), *Advancing quantitative methods in second language research*. New York: 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.

Kline, P. (1999). *The handbook of psychological testing*. London: Routledge.

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). New York, NY: Routledge.

Lazaraton, A., Riggenbach, H., & Ediger, A. (1987). Forming a discipline: Applied linguists' literacy in research methodology and statistics. *TESOL Quarterly, 21*, 263–277.

Loewen, S., Ballard, L., Crowther, D., Gonulal, T., Isbell, D. R., Lim, J., Maloney, J., & Tigchelaar, M. (under review). How statistically literate are SLA and applied linguistics researchers? Data from North America and Europe. *Studies in Second Language Acquisition*.

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). Exploratory factor analysis and principal components analysis. In Plonsky, L. (Ed), *Advancing quantitative methods in second language research*. New York: Routledge.

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.

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.

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., & 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.

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.

Roever, C., & Phakiti, A. (2018). *Quantitative methods for second language research: A problem-solving approach*. London: Routledge.

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance*, *1*(1), 15-20.

Schield, M. (2004). Statistical literacy and liberal education at Augsburg College. *Peer Review, 6,* 16-18. Retrieved from www.StatLit.org/pdf/2004SchieldAACU.pdf.

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

Thomas, M. (2013). The doctorate in second language acquisition: An institutional history. *Linguistic Approaches to Bilingualism*, *3*(4), 509-531.

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.

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

Appendix A. Statistical background questionnaire

1. Age _____

2. Gender: Male __ Female__

3a. What is your current academic position?

o MA student
o PhD student
o Other (Please specify) _____

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

3c. What is your major field of study?

o Applied Linguistics
o TESOL/TEFL
o Second Language Acquisition
o Foreign Languages
o Language Testing
o Education
o English
o 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.

Map of the United States and Canada

4. Please rate the following statements

- To What extent do you identify yourself as a researcher?

Not at all Exclusively

1 2 3 4 5 6

- To what extent do you conduct quantitative research?

Not at all Exclusively

1 2 3 4 5 6

- To what extent do you conduct qualitative research?

Not at all Exclusively

1 2 3 4 5 6

5a. Approximately how many quantitative analysis/statistic courses have you taken? ____

5b. When did you take your last quantitative analysis/statistics course? (E.g., Fall, 2014) _____

5c. Which department(s) offered the quantitative analysis/statistics course(s) that you took? (Please select all that apply)

- 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
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 Expert

1 2 3 4 5 6

Appendix B. Statistical literacy assessment for second language acquisition (SLA for SLA) survey

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.

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.

Descriptive statistics for all three tasks

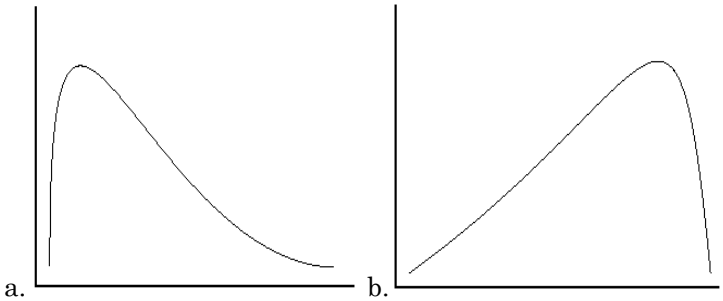
	Mean	Median	Mode	SD	95% CIs Lower bound Upper bound
Picture difference task	7.09	8	9	3.91	[5.03 - 9.15]
Consensus task	4.00	2	2	1.00	[2.36 - 4.88]
Map task	6.23	9	11	5.61	[6.17 - 10.29]

7. Which of the following statements is TRUE?

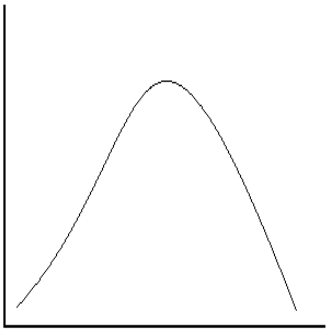
a. The variance in the map task data is the highest
b. The variance in the picture difference task data is the highest
c. The variances in the picture difference task data and the map task data are the same
d. I don't know

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.



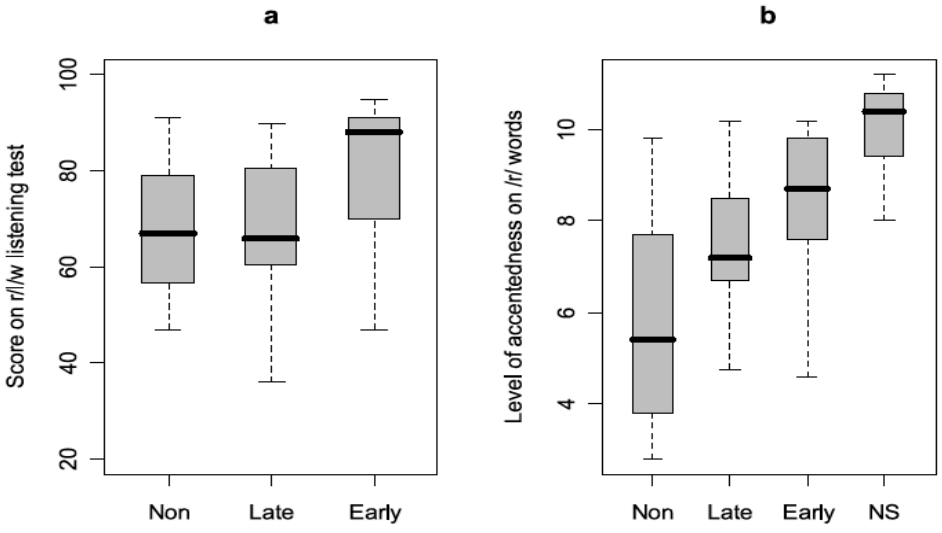
a. b.



c. d. I don't know

Graphs for map task data

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



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, 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.

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

Internet

Calculator

Other colleagues

Other_____

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!