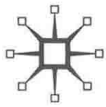
Writing about Quantitative Research in Applied Linguistics

Lindy Woodrow
University of Sydney, Australia

palgrave
macmillan



First published 2014 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN 978–0–230–36996–2 hardback
ISBN 978–0–230–36997–9 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

Writing about Quantitative Research in Applied Linguistics

Also by Lindy Woodrow

ADAPTIVE SECOND LANGUAGE LEARNING

TEACHING ACADEMIC WRITING (*joint author*)

Examples

Tables and Figures

Tables

Figures

Acknowledgements

The examples used in this book are taken from journal articles, theses and dissertations. I would like to thank the following authors for their permission to use these extracts (full bibliographic details for the extracts are provided at the end of each chapter):

Chapter 2: R. Serrano (2011) and H. Zhong (2008) for the purpose statement and research questions; S. Ryan (2008) for writing about methodological approach; J. E. Peng (2010) for extracts from our joint article; T. Tode (2003) and R. Schoonen, A. Gelderen, R. Stoel, J. Hustjin and K. Glopper (2011) for reference to time in an abstract; G. S. Levine for referring to questionnaire.

Chapter 3: N. Mills (2011) for the discussion of reliability and validity; M. K. Rai, L. C. Loschky, R. J. Harris, N. R. Peck and L. G. Cook (2011) and H. Lin, T. Chen and F. Dwyer (2006) for writing about reliability coefficients; J. Lee for the reliability and validity section; H. Rose (2010) for reference to ethics; G. S. Levine (2003) for writing about research procedure.

Chapter 4: S. Graham and E. Macaro (2008) for describing a stratified sampling procedure; I. Purfahl and N. Rhodes (2011) for sampling criteria for stratified sampling; C. B. Zimmerman (1997) for writing about participants; W. Tsou (2011) for using a convenience sample; J. E. Peng (2009) for an introductory paragraph about participants; F. Han (2008) for writing about participants.

Chapter 5: M. Papi and Y. Teimouri (2012) for presenting descriptive statistics; K. Kondo-Brown (2004) for Table 1 (Example 5.1); M. S. Uggen (2012) for Table 3 (Example 5.4).

Chapter 6: R. Gardner, A.-M. Masgoret and J. Tennant (2004) for using *t*-tests to screen data; S. Webb, J. Newton and A. Cheng for using paired *t*-tests; C. Taylor and R. Lafayette (2010) for using *t*-tests as a post-test.

Chapter 7: A. van Gelderan, R. Oostdan and E. van Schooten (2010) for reporting ANOVA; M. Lucas (2011) for writing about three factorial ANOVA; M. Takimoto (2008) for writing about two-stage ANOVAs; L. Vandergrift and M. Tafaghodtari (2010) for writing about ANCOVA in an experiment.

Chapter 8: S. Matsumara and G. Hann (2004) for reference to regression, F. Han (2008) for reference to regression in a proposal;

R. L. Cochran and R. S. McCallum (2010) for text and Table 3 (Example 8.3); H. Cheung, M. Chan and K. Chong (2007) for text and Table 3 (Example 8.4).

Chapter 9: D. L. Nisbet, E. R. Tindal and A. A. Arroyo (2005) for Table 2 (Example 9.2); Y. Mori, K. Sato and H. Shimizu (2007) for Table 4 (Example 9.3); S. Ryan (2008) for text and Table 7.3 (Example 9.4); J. Ong and L. Zhang (2012) for correlation analysis for inter-rater reliability.

Chapter 10: L. Vandergrift, C. C. M. Goh and C. J. Mareschal (2006) for exploratory and confirmatory factor analysis in an abstract; R. L. Sparks, J. Patton, L. Ganschow and N. Humbach (2011) for reference to exploratory factor analysis in the results section; P. Winke (2011) for text and Table 2 (Example 10.3); Language Learning for permission to reproduce Figure 1 as a factor diagram and text from L. Woodrow (2006); J. E. Peng (2009) for text and Figure 5.1 (Example 10.5).

Chapter 11: J. F. Chen, C. A. Warden and H.-T. Chang (2005) for research questions and Figure 1 (Example 11.1); S. Andringa, N. Olsthoorn, C. van Beuningen, R. Schoonen and J. Hulstijn (2012) for the steps in SEM analysis; H.-C. Yang (2012) for the assumptions and fit indices; K. Fushino (2010) for Figure 2 (Example 11.4) and model estimation; J. E. Peng (2009) for Table 5.1 (Example 11.5); J. E. Peng (2010) joint article for Table 3 (Example 11.6).

Chapter 12: J. M. Dewaele, K. V. Petrides and A. Furnham (2008) for testing to see whether parametric tests are appropriate; H. Zhong (2008) for justification for the use of non-parametric tests taken from a Master's dissertation; K. Barkaoui (2010) for Mann–Whitney test of significant differences; Y. Yilmaz (2011) for Wilcoxon signed-ranks test; J. McEwan-Davis (2007) for Tables 5 and 6 (Example 12.5); S. Li (2012) for Friedman's ANOVA and Kruskal–Wallis in results section.

The University of Sydney Papers in TESOL for extracts from L. J. Woodrow (2006a). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney Papers in TESOL, 1*, 51–70.

I wish to thank the editorial team at Palgrave Macmillan, in particular Olivia Middleton, Libby Forrest and Nicola Jones for their support and guidance during the writing of this book.

Glossary of Key Terms

ANCOVA	This acronym refers to analysis of covariance. It is similar to ANOVA in that it is used to compare the difference in means of two or more sets of scores. However, in addition to independent and dependent variables ANCOVA controls for an additional variable(s) or covariate(s).
ANOVA	This is the acronym used for analysis of variance. ANOVA is a statistical method for investigating the difference in means of two or more sets of scores. A one-way ANOVA has one dependent variable and one independent variable. An independent-measures ANOVA uses two independent groups while a repeated-measures ANOVA uses scores from one group on two or more occasions. A factorial ANOVA is used to analyse different levels of the independent variable.
Central tendency	This refers to the range of scores on a questionnaire or a test and how these are dispersed. The mean (*M*), median (*Mdn*) and mode (*Mode*) are all measures of central tendency.
Chi-squared (χ^2)	This refers to statistical analysis. There are two types: the goodness-of-fit chi-squared test and the chi-squared test for independence. The goodness-of-fit chi-squared test is used to compare expected and actual outcomes. The chi-squared test for independence explores the relationship between two categorical variables.
Cohen's *D*	This is the statistic that can be used to calculate effect size for *t*-tests.
Confirmatory factor analysis (CFA)	This belongs to the factor analysis family. It is a statistical technique used to confirm relationships between variables using structural equation modelling techniques.

Convenience sample	This is a sampling procedure whereby the sample participants are drawn from extant groups, for example a class of students.
Correlation	This is a statistical technique used to investigate the relationship between variables. The correlation coefficient (r) indicates the strength of this relationship. Pearson's product–moment correlation is used with parametric tests and Spearman's rho (ρ) for non-parametric tests. Partial correlation is an analysis that examines the correlation between two variables controlling for a third.
Correlation matrix	This is the table used to report correlations between variables. It includes all the correlation coefficients and the significance of these in a triangular half table.
Cross-sectional research design	This refers to a research design that collects data on one occasion only. A cross-sectional project is typically a one-off questionnaire given to a large sample and provides a snapshot of the variables of interest.
Descriptive statistics	This refers to data collected about participants that involve no inference, for example age, gender, and home language. Included in descriptive statistics will be measures of central tendency, for example mean, median, mode and standard deviation.
Dependent variable (DV)	A variable is something that can change. In a cause–effect relationship, the dependent variable is the measure that is thought to be influenced by, that is the effect of, the independent variable.
Digital Object Identifier	This is the unique alpha-numeric code given to a document that does not change over time. Most recent publications have these; some publications and thesis examiners insist on an author providing these.
Dichotomous variable	This is a variable that has two possible responses and is usually represented by a yes/no question.
Effect sizes	These provide information about the magnitude of results of statistical tests. Most journals require these to be reported.

Eta squared (η^2)	This is a statistic that can be used to calculate effect sizes for *t*-tests.
Experimental group	This refers to the group of participants that receives a treatment or intervention. This is contrasted with the control group that does not receive this. Usually the two groups are then compared to evaluate the effect of the treatment or intervention.
Experiment	For a research project to be classed as a true experiment, the participants need to be randomly selected.
Exploratory factor analysis (EFA)	This is a factor analysis technique that explores relationships based on correlations between variables to indicate underlying groupings. It is often used to provide evidence for the validity of a questionnaire.
Factor analysis	This is a series of statistical techniques that can be exploratory (EFA) or confirmatory (CFA). It is used to accept or reject hypotheses about relationships between a series of variables.
Fit indices	In structural equation modelling, a series of statistical analyses are conducted to evaluate the fit of the proposed model and the data. These are known as fit indices and determine whether the model is accepted or rejected.
Friedman ANOVA	This is the statistical test that is the non-parametric equivalent of repeated measures ANOVA. It is used to identify differences on two or more measures of one group of participants.
Hypothesised model	This is the term used in structural equation modelling for the initial model of relationships between variable. This is then tested and modified as necessary.
Hypothesis	This is a statement that is accepted or rejected by the research. A research project may use hypotheses or research questions.
Independent variable (IV)	A variable is something that can change. In a cause–effect relationship the independent variable is the measure that is thought to influence (cause) the dependent variable.

IELTS	The International English Language Testing Service test is a high-impact test widely used to assess English proficiency for university entry.
Independence of means	This means that the scores from one person must not influence the scores of another. This is usually tested statistically.
Inferential statistics	By using certain statistical tests, with a sufficiently large and randomly selected sample, researchers can make generalisations about the results of the research to a wider group or population of similar people.
Inter-rater reliability	This refers to the similarity in scoring of two or more raters. This is usually done statistically with an established minimum level, for example $r = 90$.
Instrumentation	This refers to the questionnaire or data collection tests used in a research project.
Informed consent	This refers to ethics in research conduct. Participants need to be informed about the research project before they agree to take part. Inherent in this is voluntary participation.
ISI	This is the acronym for the Institute of Science Index. Journals that are included in this list are highly regarded.
Kolmogorov–Smirnov (K–S) test	This is a statistical test of normality often used to decide whether to continue using parametric techniques or to change to non-parametric tests.
Kruskal–Wallis test	This is a statistical test that is the non-parametric equivalent of a one-way ANOVA. It is used to identify differences between two or more groups.
Latent variable	A variable that is unobserved and is often referred to as a latent construct. A latent variable is hypothesised and supported by evidence from relationships between the observed variables (items).
Likert scale	A type of questionnaire that has responses measured in equal parts. For example, a five-point Likert scale may be categorised as (1) strongly agree, (2) agree, (3) neutral, (4) disagree and (5) strongly disagree.

Longitudinal research design	This refers to a data-collection technique that has multiple rounds. Participants typically provide data on a number of occasions over a period of time. It is used to examine the development of a specified variable, for example language learning.
Mann–Whitney test	This is a statistical test that is the non-parametric equivalent of an independent-samples *t*-test. It is used to compare the scores of two groups.
MANOVA	Multivariate analysis of variance (MANOVA) is a statistical analysis that is used to compare the difference in means of sets of scores where there is more than one dependent variable, rather than just one as in ANOVA.
Mean (*M*)	This is a measure of central tendency used to refer to the average. This is obtained by adding all the scores together and dividing them by the number of participants.
Measurement model	In structural equation modelling, this reflects the observed variables and includes the confirmatory factor analyses.
Median (*Mdn*)	This is a measure of central tendency, referring to the exact middle score of a range of scores.
Mode (*Mode*)	This is a measure of central tendency that shows the most frequently reported score of the group of participants.
Multivariate	This describes an analysis that includes multiple variables.
Multicollinearity	This term is used when the variables in a study are highly correlated.
Multiple regression	This is an analysis that can include multiple independent variables and predicts scores on a dependent variable.
Non-parametric tests	These are used when data have violated the assumptions of parametric tests, for example normality.
Observed variable	This term is used in some research techniques, such as factor analysis and structural equation modelling, to refer to the measured variables or items of a questionnaire or test.

Operationalisation	This reflects how the constructs or variables in a research project are measured, for example by questionnaire items.
Outliers	These are scores that are distant from the mean and can distort the data set. Outliers can be univariate or multivariate.
Path diagram	This is the visual representation of a structural model of variables and their relationships.
Parametric tests	These are the usual statistical techniques that adhere to assumptions of normality and random sampling.
Population	This term refers to the larger group from which a sample is drawn. A quantitative research project usually aims to relate findings from research using a sample to the larger population through inferential research techniques.
Principal Components Analysis (PCA)	This is a type of factor analysis where variables are reduced to a smaller set of variables with all the variance used.
Quasi-experiment	This is used with naturally occurring groups rather than a random sample as dictated by a true experiment.
Reliability	This refers to the consistency of results on an instrument such as a language test or a questionnaire. The most commonly referred to method is Cronbach's alpha (α).
Regression	This is a statistical technique that predicts scores on a dependent variable (DV) based on one or more independent variables (IV). Linear regression reflects a simple analysis between two or more variables. Sequential regression is used when the variables are entered into the analysis in an order specified by the researcher. Stepwise regression refers to an analysis where the variables are entered in an order specified by the computer software.
Research questions	These are the questions reflecting the purpose of the research that are stated and answered explicitly in a quantitative research text.