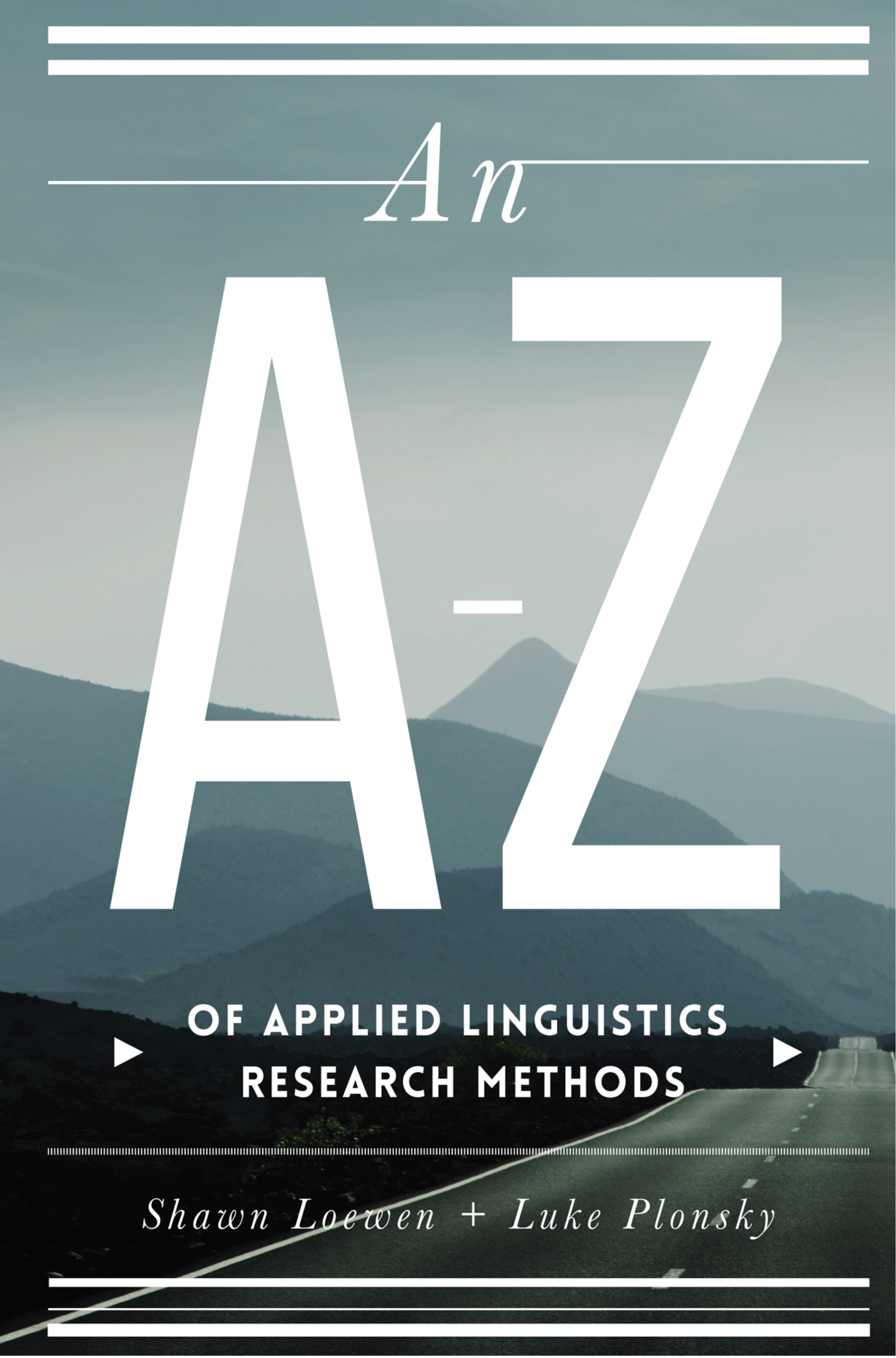
An
A-Z
OF APPLIED LINGUISTICS
RESEARCH METHODS
Shawn Loewen + Luke Plonsky

AN A – Z OF APPLIED LINGUISTICS RESEARCH METHODS

An A – Z of Applied Linguistics Research Methods

Shawn Loewen and Luke Plonsky

macmillan education palgrave

First published 2016 by
PALGRAVE

Palgrave in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of 4 Crinan Street, London, N1 9XW.

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

Palgrave is the global imprint of the above companies and is represented throughout the world.

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

ISBN 978-1-137-40321-6 ISBN 978-1-137-40322-3 (eBook)

DOI 10.1007/978-1-137-40322-3

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.

Loewen, Shawn, author.

An A – Z of applied linguistics research methods / Shawn Loewen and Luke Plonsky.

pages cm

Includes index.

Summary: "A brief and accessible introduction to the concepts and techniques used in applied linguistics research, which will be illustrated using real-life examples. The book covers both qualitative and quantitative research design, sampling procedures, instrumentation and analyses found in applied linguistics research" — Provided by publisher.

1. Applied linguistics—Research—Handbooks, manuals, etc. 2. Applied linguistics—Resesarch—Methodology—Handbooks, manuals, etc. I. Plonsky, Luke, author. II. Title.

P129.L64 2015

418.0072'1—dc23

2015033257

Shawn:

To Treva, who taught me how to write.

Luke:

To my father, Matthew Plonsky, and grandfather, Andrew Plonsky, who passed on to me a love of numbers.

Contents

Acknowledgments

We are indebted to numerous individuals who have helped us as we have written this book. First and foremost we would like to thank our editors, Paul Stevens and Aléta Bezuidenhout, for their encouragement and support throughout the entire process. This book would simply not have been possible without them. Next, enormous thanks go to Magda Tigchelaar and Dan Brown, who worked as our research assistants and helped us with all the things that professors do poorly, such as chasing up examples for entries, securing copyright permissions, and creating an index. Special mention also goes to Talip Gonulal for his helpful comments on drafts of the manuscript. We would also like to thank Cambridge University Press for their enlightened copyright policy which made it possible for us to include numerous graphics without going into debt. Figure 34: "Facets map resulting from a Rasch Analysis" is reproduced courtesy of © R Foundation: http://www.r-project.org. Figure 39: "Data view screen in SPSS" is reproduced courtesy of International Business Machines Corporation, © International Business Machines Corporation. Finally, we would like to thank our long-suffering families for their support throughout this process. To Pamela Loewen for the sacrifices she made, and to the Loewen kids—Winona, Patrick, and Austin—who dread to hear the words "Dad is working on a book deadline." I couldn't do it without the love, support, encouragement, and distraction that I get from all of you! Big thanks to Pamela Plonsky as well! And I (Luke) thank Mateo, Ruby, and Rose, precisely because they have no idea about research methods or book deadlines. You keep me "normal."

Introduction and User Guide

Conducting research is not always easy. Each study involves making numerous choices, many of which are not inherently good or bad but rather, represent a unique theoretical or methodological stance, and present several advantages and disadvantages. After we decide what phenomenon we will study (no small feat in and of itself), we have to make a host of decisions about our sample, design, data collection instruments, procedures, analyses, and interpretations. And, of course, each of these major decision points leads to additional considerations and choices, all requiring an understanding of their corresponding strengths and weaknesses relative to numerous possible alternatives.

Further complicating the researcher's life are the seemingly endless terms for the concepts and techniques that must be learned and applied. Researchers in applied linguistics must distinguish, for example, between terms such as nominal and interval data, Type I and Type II errors, reliability and validity, and case studies and ethnographies, just to name a few. We understand this challenge, and we're here to help!

Our goal in writing this book is to enable researchers, reviewers, consumers of research, and students of research methods in applied linguistics to quickly access an overview or refresher on critical terms and the methodological practices and issues associated with them. We also tried to do so in a format that can be taken advantage of quickly, almost at a glance. Traditional methodological texts necessarily describe key concepts and terminology embedded deep within larger discussions. The entries in this book—more substantial than a glossary but not as lengthy as an encyclopedia entry—give the reader "bite-sized" overviews of key terms, explaining their background, meaning, purpose, and place in the field. We also recognize, however, that many of the concepts described here are related. It's difficult to talk about parametric statistics, for example, without some understanding of a normal distribution. Likewise, it makes little sense to consider epistemology in isolation from ontology. We have therefore emboldened terms with entries found elsewhere in the book; we've done this in an effort to preserve and alert readers to the interrelatedness of terms and concepts. It's very likely, in fact, that skipping around from entry to entry will be the most useful way to use this book. We don't actually envision that anyone would use this as their stand-alone research methods textbook, nor would most people read the book cover to cover. Rather, we recommend the use of this book as an accompaniment—a sidekick, if you will—to whatever text or texts readers use as their go-to, comprehensive guide to research methods.

In addition to explaining a bit about our purpose and approach in writing this book, we also wanted to share a few tips that might make it more useful as a reference and research companion. For one, the text provides numerous examples from real applied linguistics research in order to provide context to the terms and concepts. We use these to illustrate and exemplify the sometimes abstract concepts we cover. References to such studies and to further reading are also provided, along with their relevant entries, to allow readers to track them down for additional information. Another aid we've included are visuals, lots of them. If you haven't done so yet, take a moment to flip through the book. Many of the concepts we cover here lend themselves very nicely to graphic displays, and we've tried to take advantage of this as often as possible as a kind of pedagogical tool.

Another issue that comes up in reading a text on research methods (or any kind of text, really) is the target audience. That is, who is this book written for? In short, everyone. We'll explain. First of all, we tried to keep a number of diverse audiences in mind while writing this book. We especially sought to consider the needs of students and novice researchers. But this book is not only for those new to the field or to research. This book is for anyone who has ever needed a quick reminder on the difference between construct validity and face validity. This book is for anyone looking for an example of a longitudinal study. This book is for anyone wanting a quick review of the different types of reliability. In other words, this book is for anyone who conducts, reads, or reviews research in applied linguistics. Meanwhile, we are well aware of the challenges in preparing a text for an audience that is potentially very diverse in terms of training, background, research interests, theoretical approaches, and so forth. We also admit that there may be moments in the book when we have presented what some would consider to be an oversimplification of certain concepts and/or techniques for the sake of clarity. At the same time, despite our efforts to keep the novice researcher in mind at all times, it is also possible that certain entries may appear overly technical or complex. If you find yourself in this situation, please don't lose heart! Everyone feels this way at some point or other regarding research methods.

Despite these and many other challenges, an understanding of research methods is critical to our field. It's actually difficult to overstate the importance of methodological knowledge. If applied linguistics researchers—individually or collectively—lack conceptual or technical know-how, the studies we produce cannot accurately inform theory or practice. Furthermore, it is critical that those of us who produce research possess a keen understanding of methodological concepts, practices, and so forth, since trusted gatekeepers, journal reviewers, and editors need to be able to assess the value of the thousands of reports submitted each year to applied linguistics journals. Likewise, consumers and practitioners, such as teachers, test developers, and language program administrators, must be able to evaluate studies relevant to their work as a means to know what and how they can be applied to their own contexts. In other words, all of us within the applied linguistics community must possess some level of methodological knowledge.

We have tried to make this book as comprehensive as possible, including general research terms from α to *z- score*. In addition, we have focused on discipline-specific topics from *action research* to *VARBRUL* for applied linguists. However, without reading too many entries, it will become clear that our own specialized area of research within applied linguistics pertains to second language acquisition (SLA). Many of the studies we have chosen to illustrate specific terms are from SLA, and even from our own research. The reason for this is primarily a practical one. These are the studies that are most familiar to us. However, the terms in this book were chosen specifically for their broader appeal, and we do not regard SLA research as more valuable than other areas within applied linguistics.

Also, it is our aim to encourage good research practices, both in conducting and reporting research. That being said, not all published articles contain all of the information that we might hope for. In considering the examples to include in the entries, we sometimes found studies that were good, but did not report everything that we might recommend. In such cases, we used the example as-is, but we would ask readers to be aware that, consequently, there may be a discrepancy between what we recommend reporting compared to what we include in the real-life examples.

In the end, we hope you will find this volume useful for you, whatever your level of research experience. In addition, we wish you many happy years of research design, data collection, analysis, and reporting!

A a

α (see Alpha)

A priori test (see Pretest)

Action research

Research generally conducted by teachers or teacher trainees in order to investigate a **research question** that is relevant to their particular pedagogical context. Action research often involves the examination of a practical problem or issue in the classroom. In many cases, action research is an iterative process in which an initial question or **hypothesis** is proposed, and an effort is made to address the question. Based on the results of the investigation, a new hypothesis might be put forward and investigated. An example of a collaborative action research project is presented by Banegas, Pavese, Velázquez, and Vélez (2013), who examined the integration of content and language learning in their own English as a foreign language classes using a three cycle process. In cycle 1, the researchers identified specific issues related to content and language integrated learning (CLIL) in their context, and subsequently designed, implemented, and evaluated teaching materials. Cycles 2 and 3 involved similar action, intervention, and evaluation activities based on information gained from previous cycles. Banegas et al. reported on the positive impact of action research for their own professional development, and the resulting benefits for their students.

In many cases, the information gained from action research is intended primarily for the teachers conducting the research and not for a wider audience. Because action research addresses questions that are of interest to specific teachers in specific contexts, there is often not a concern with generalizing the results to other settings. Action research is sometimes discounted because it is not viewed as being as rigorous or systematic as **experimental**, **quasi-experimental**, or other types of research; nevertheless, it can be quite informative, and it has become an important component of applied linguistics research.

Banegas, D., Pavese, A., Velázquez, A., & Vélez, S. (2013). Teacher professional development through collaborative action research: Impact on foreign English-language teaching and learning. *Educational Action Research, 21*, 185–201.

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. Abingdon: Routledge.

McDonough, K. (2006). Action research and the professional development of graduate teaching assistants. *Modern Language Journal, 90,* 33–47.
Yuan, R., & Lee, I. (2015). Action research facilitated by university-school collaboration. *ELT Journal, 69,* 1–10.

Adjusted R^2

A statistical value, based on $\boldsymbol{R^2}$, that indicates how well a regression model generalizes from a **sample** to an entire **population**. When conducting a regression, researchers want to determine how well certain variables (called predictor variables) predict scores on another variable (called an outcome variable). For example, Venkatagiri and Levis (2007) wanted to know how closely L2 speakers' phonological awareness was associated with their comprehensibility when speaking. Their **linear regression** analysis resulted in an R^2 value of .241, indicating that 24% of the **variance** in speaker comprehensibility could be accounted for by the **predictor variable** of phonological awareness. The R^2 value applies only to the group being investigated in the research study; consequently, in order to generalize to the larger population, an adjusted R^2 value was calculated. Venkatagiri and Levis reported an adjusted R^2 of .190, leading the authors to claim that, conservatively, phonological awareness accounts for 19% of the variance in comprehensibility scores. The adjusted R^2 value indicates the loss of predictive power that occurs if the model were applied to the larger population. The less difference there is between the R^2 and the adjusted R^2 values, the better the model.

Venkatagiri, H. S., & Levis, J. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness, 16*, 263–277.

Alpha (level)

A numeric cutoff point for determining if a statistical test has produced a significant result. Often represented by the Greek letter $\boldsymbol{\alpha}$, the alpha level is typically set at .05, meaning that there is a less than 5% probability that the result of the statistical test is interpreted as significant when it should not be, and that a **Type I error** has been committed. The alpha level desired by a researcher should be determined at the beginning of statistical analysis, based on the level of statistical certainty required for the research study. In more exploratory, low-stakes research, an alpha level of .10 or .20 is considered acceptable. For instance, Ammar and Spada (2006) set their cutoff for statistical significance in comparing **pretest** results at .10. This choice was motivated by an interest in ensuring that any pretreatment differences between groups would be detected. With the more traditional but stricter alpha of .05, they might have missed group differences that ought to be taken into account in the **posttest** analyses (see Plonsky, 2015a, 2015b). Such a mistake, a **Type II error**, would pose a threat to the **validity** of the findings. For the same reason, in medical research, a much stricter alpha level of .001 is often set, due to the need for precise results.

The ***p*-value** produced by the statistical test is compared to the alpha level to determine if a test result should be considered statistically significant. In **Null Hypothesis Significance Testing (NHST)**, the alpha level is often treated as invariant and dichotomous, with all values greater than .05 being viewed as non-significant and all values less than .05 being significant; however, there is much criticism of such a categorical use of the alpha level because according to such a strict criterion, the values of .06 and .96 are both treated equally, as non-significant, even though they express a 6% as opposed to 96% probability of obtaining the observed data or effect, assuming such an effect does not exist in the **population**. Consequently, there has been a call for caution in using NHST, with suggestions to use additional values, such as **effect sizes** and **confidence intervals**, to assess the importance and stability of a finding (e.g., Norris, 2015; Plonsky, 2015b).

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*, 543, 574.

Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions for reform. *Language Learning, 65*(Supp. 1), 97–126.

Plonsky, L. (2015a). Quantitative considerations for improving replicability in CALL and applied linguistics. *CALICO Journal, 32,* 232–244.

Plonsky, L. (2015b). Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23–45). New York: Routledge.

Analysis of covariance (ANCOVA)

A type of **analysis of variance** that attempts to account for the effects of a moderating variable on a **dependent variable**. The goal of ANCOVA is, therefore, to arrive at a more accurate estimate of the effect of the **independent variable**(s) on the dependent variable(s), without the influence of other intervening variables. One of the ways that ANCOVA is sometimes used in applied linguistics research is in **quasi-experimental designs** investigating the effects of an instructional treatment by including a **pretest** and at least one **posttest**. Ideally, all groups should score similarly on the pretest in order to ensure that any gains made on the posttest are due to the treatment and not to other variables, such as initial differences in proficiency levels between the groups. However, if there are differences between the groups on the pretest, such as may occur when using **intact classes**, then the pretest scores can be treated as a **covariate** in the ANCOVA, and the analysis will examine the effects of the treatment after the moderating variable of proficiency level has been accounted for. In this way, researchers can obtain a clearer picture of the effects of the instructional treatment. Ammar and Spada (2006), for example, were interested in comparing the effects of two types of feedback—recasts and prompts—on the development of English third-person possessive determiners *his* and *her*. The pretests revealed, however, that there were proficiency differences between the groups at the beginning of the study. Therefore, the authors used ANCOVA, instead of ANOVA, to compare the two treatment groups and the **comparison group**. The use of ANCOVA is somewhat controversial, with some people suggesting that the effects of a covariate can never really be partialled out (Field, 2013). Nevertheless, ANCOVA is still sometimes used in applied linguistics research. Assumptions to be investigated before conducting an ANCOVA include: (a) homogeneity of regression slopes, and (b) independence of the covariate and treatment effect (Field, 2013).

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L 2 learning. *Studies in Second Language Acquisition, 28*, 543, 574.
Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Thousand Oaks, CA: Sage.

Analysis of variance (ANOVA)

A statistical procedure that compares the differences in the means of three or more groups. ANOVA is the most commonly used statistical test in applied linguistics research (Gass, 2009; Plonsky & Gass, 2011). By enabling the researcher to compare the means of more than two groups, ANOVA expands on the ***t*-test**, which compares the means of just two groups.

There are several different types of ANOVAs, and the naming conventions used to refer to them are sometimes confusing and inconsistent (see Table 1, which contains example studies). However, one thing that all ANOVAs have in common is that they allow only one **dependent variable** measured on a **continuous scale**, while **independent variables**, of which there may be multiple, are **categorical** in nature. Another important distinction between different types of ANOVAs involves determining whether the data being analyzed come from different groups (**between-groups**) or the same group (**within-groups**). Below are described several different types of ANOVAs, beginning with the simplest and continuing with more complex ones:

- A one-way analysis of variance has one dependent variable and one independent variable with three or more different levels or groups. This test compares the means of those groups to see if there are statistically significant differences between the groups. In order to interpret a statistically significant one-way ANOVA, a **post hoc** analysis must be conducted in order to determine the exact nature and location of the differences. Because there are three groups, it might be the case that only two groups differ significantly from each other, or it might be the case that each group differs significantly from the others. A post hoc test (e.g., LSD, Tukey's HSD, Scheffe) will reveal these differences.

Table 1 Examples of different types of ANOVAs in SLA

Study	**Type of ANOVA (reported)**	**Type of ANOVA (as described here)**	**Dependent variable**	**Between-groups variable(s)**	**Within-groups variable(s)**
Lyster (2004)	Analysis of variance	Mixed design ANOVA	Test scores	Group (FFI-prompt, FFI-recast, FFI-only, Comparison)	Time (Pretest, Posttest 1, Posttest 2)
Riazantseva (2012)	Repeated measures ANOVA	Repeated measures ANOVA	Accuracy rate	None	Outcome measure (in-class essay, in-class summary, at-home summary)
Saito (2013)	Three-factor ANOVA	Mixed design ANOVA	Test scores	Treatment group (FFI+EI, FFI only, Control)	1. Test (controlled versus spontaneous production) 2. Test Time (pretest versus posttest)
Sheen (2008)	One-way ANOVA	One-way ANOVA	Anxiety scores	Group (high anxiety-recast, low anxiety-recast, high anxiety-control, low anxiety-control)	None

Note: Tukey's pairwise comparisons were carried out as post hoc tests in each of the studies except for Riazantseva, who did not conduct post hoc tests.

- A factorial ANOVA includes two or more independent variables, and each variable contains different groups, meaning that it is an independent, rather than repeated measures, ANOVA. Sometimes when referring to ANOVAs, the number of independent variables is indicated in the name. For example, a two-way ANOVA has two independent variables, while a three-way ANOVA has three independent variables, and so on. Just to add to the confusion, sometimes researchers use the number of levels in each independent variable to name the ANOVA. So, a 2 × 3 ANOVA has two independent variables, one with two levels and one with three levels of measurement. A 2 × 3 × 3 ANOVA has three independent variables, with one measured with two levels and the others with three levels. A factorial ANOVA will produce **main effects** for each independent variable, as well as **interaction effects** that compare each variable against the others. With two independent variables, there will be two main effects (X, Y) and one interaction effect (X × Y), while with three independent variables there will be three main effects (X, Y, Z), three two-way interaction effects (X × Y, X × Z, Y × Z), and one three-way interaction effect (X × Y × Z). Three-way interaction effects and greater can be difficult to interpret; consequently, other types of analysis such as **multiple regression** may be preferable with three or more independent variables. As with one-way ANOVAs, post hoc tests should be conducted following statistically significant factorial ANOVAs in order to determine the exact location and direction of the differences.
- Repeated measures ANOVAs compare three or more independent variables (e.g., tests) taken from the same group. As such, it is similar to a paired-samples *t*-test. For example, a repeated measures ANOVA might compare the performance of one group on three different tasks, or at three different times (such as a **pretest**, **posttest**, and **delayed posttest**).
- Mixed design ANOVAs have at least two independent variables, one or more of which is between-groups and one or more of which is within-groups. This is a common design in SLA research when a researcher compares two or more treatment groups on a pretest and one or more posttests in order to gauge the effectiveness of some type of instructional treatment.

Each of these ANOVA types has specific **assumptions** that should be met before conducting the analysis. The two primary assumptions for the between-groups ANOVAs are a **normal distribution** of data and **homogeneity of variance**. A main assumption for repeated measures ANOVAs is **sphericity**, which is conceptually similar to homogeneity of variance.

In addition to running post hoc analyses, another good way of visualizing the relationships between the groups is through graphs, especially **line graphs**, which provide evidence of interaction effects when lines are distinctly non-parallel.

In addition to indicating statistical differences between groups, it is important to note the **eta-squared** value associated with an ANOVA. This value is an **effect size** for ANOVA and indicates the amount of variance in the dependent variable that can be accounted for by the **categorical variable**(s) under investigation. Also, because post hoc analyses for ANOVA involve comparing individual pairs of mean scores, the **Cohen's *d*** effect size is generally more appropriate at this phase of the analysis.

Gass, S. (2009). A survey of SLA research. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 3–28). Bingley: Emerald.

Lyster, R. (2004). Differential effects of prompts and recast in form-focused instruction. *Studies in Second Language Acquisition, 26,* 399–432.

Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, *61*, 325–366.

Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System, 40*, 421–430.

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*(1), 1–29.

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning, 58*, 835–874.

ANCOVA (see Analysis of covariance)

ANOVA (see Analysis of variance)

Aptitude-treatment interaction (ATI)

Research that attempts to investigate the ways in which the effects of different types of instruction vary according to individual learner differences, such as general learning abilities, learning styles, and language learning ability. The goals of ATI research are both theoretical and practical. From a theoretical perspective, research in this domain can help inform the interaction between learner-external and learner-internal variables. More practically speaking, one of the goals of ATI research is to match specific types of instruction with learners who will maximally benefit from that type of instruction based on specific individual difference characteristics. Interest in ATI research is often traced back to educational psychology, where over half a century ago Cronbach (1957) called for greater collaboration between researchers examining causal relationships (e.g., instructional effects) and those interested in correlation relationships (e.g., individual differences). In applied linguistics, interest in ATI research has been gaining some momentum. Yilmaz (2013), for example, examined working memory and language analytic ability as moderators of the effects of two types of corrective feedback, namely explicit correction and recasts. Among other results, Yilmaz found that explicit feedback was more effective than recasts only among participants with higher working memory and language analytic ability. In a comparable design, Li (2013) also found an interaction between the effects of implicit feedback and learners' language analytic ability, suggesting that higher analytic ability allowed learners to benefit from implicit feedback.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

DeKeyser, R. (2009, October). *Variable interaction in SLA: Much more than a nuisance*. Plenary address given at the Second Language Research Forum, East Lansing, MI.

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *Modern Language Journal, 97*, 634–654.

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics, 34*, 344–368.

Assumption

Conditions that must be met before conducting specific statistical analyses. All statistical tests require that the data being analyzed meet certain criteria. Some assumptions are related to the types of variables involved, whether **categorical**, **ordinal**, or **continuous**. For example, *t*-tests require a continuous **dependent variable**, such as a test score, and a categorical **independent variable**, such as treatment condition. In contrast, **chi-square tests** allow only **categorical variables**. Another type of assumption is related to the nature of the distribution of the data. For instance, **parametric statistics**, such as **ANOVAs**, require a **normal distribution** of data, while correlations and regression analyses assume both normal distributions within variables and linear relationships between them. Other assumptions about relationships among variables include: (a) **homogeneity of variance**, which assumes that the distribution of data scores is similar for groups being compared in *t*-tests or ANOVAs; (b) **sphericity**, which is similar to homogeneity of variance for repeated measures data; and (c) lack of **multicollinearity** in correlations and regressions, meaning that pairs of variables are not highly correlated. Additionally, many statistical tests assume that the data are independent, meaning that one participant's behavior does not affect the data from other participants. Finally, more advanced statistics such as **multiple regression** and **factor analysis** have additional, unique assumptions, such as homoscedasticity (residuals with equal variance) or large sample sizes.

Testing assumptions before conducting statistical analyses is an important, although often overlooked, process. If the assumptions of a specific statistical test are not met, then the results from the analysis may be unreliable

A

or invalid, thereby causing misleading interpretations of the data. Several assumptions, such as the type of variable (e.g., categorical or continuous) and **independence** of data, can be addressed in the design of the study by choosing appropriate **instruments** and participants. Other assumptions must be assessed after the data have been collected. For example, the assumption of normal distribution of data can be checked both through visual inspection of graphic representations of the data such as **histograms** and **boxplots**, as well as through statistical tests such as the **Kolmogorov–Smirnov** or **Shapiro–Wilk** tests. Similarly, homogeneity of variance can be assessed by inspecting boxplots or conducting **Levene's test** of homogeneity of variance.

When one or more assumption is violated, there are often adjustments that can be made to the data or alternative statistical procedures that can be performed. One potential modification when comparing group means that violate the assumption of normal distribution is to apply transformational procedures, such as log or square root transformations, in which a mathematical formula is applied equally to the data. Alternatively, researchers can substitute parametric tests such as *t*-tests or ANOVAs with non-parametric tests such as the **Mann–Whitney *U* test** or a **Kruskal–Wallis test**, respectively, which do not assume a normal distribution. Additionally, in some cases, statistical corrections are available when assumptions have been violated. For example, **SPSS** provides two sets of scores for independent samples *t*-tests, one assuming equal variances and the other not assuming equal variances. Likewise, when the assumption of sphericity is violated in repeated measures ANOVAs, researchers can use one of several corrections, with the Greenhouse–Geisser correction being the most common.

Finally, despite the importance of checking statistical assumptions, L2 researchers often fail to do so, or to at least report having done so (Plonsky, 2013). Nevertheless, an increase in assumption testing prior to performing statistical analyses will increase **study quality** and consequently the **reliability** and **validity** of the inferences made from **quantitative research**.

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.

A

ATI (see Aptitude-treatment interaction)

Attrition

The decrease in the number of participants over the course of a research study, due primarily to circumstances beyond the researcher's control. Although attrition may occur in **qualitative research**, it is often more critical in **quantitative research** in which researchers need specific sample sizes to perform certain statistical tests. In particular, classroom research can be susceptible to attrition due to variation in student attendance. In addition, longitudinal quasi-experimental studies involving a **pretest**, multiple treatment sessions, and several **posttests** are especially affected by the loss of research participants, because it is generally necessary for participants to take part in all components of the study. Consequently, participants may have to be excluded from the entire study, even if they miss only one or two elements, because such loss of data can make it difficult to make comparisons across all of the datapoints. This type of exclusion is sometimes referred to as listwise or casewise deletion. However, some statistical analyses can be conducted even if participants have missed part of the study, in which case researchers may wish to use pairwise deletion, which makes use of the participant's partial data. For example, if several participants missed a **delayed posttest**, but were present for the other components of the study, it is possible to conduct separate analyses with the full and partial samples. In other cases, the attrition can be quite prohibitive. The attrition rate by the time of the delayed posttest in Sanz and Morgan-Short (2004), for example, was so high that the authors chose not to include the data from the delayed posttests in their analyses. Another example of participant attrition is from Morgan-Short, Heil, Botero-Moriarty, and Ebert's (2012) study of attention allocation and think-aloud research methods. The researchers started with 410 university Spanish students; however, 45 did not reach a minimum threshold of attention allocation, and 4 participants did not follow the instructions properly. Consequently, only 361 were included in the final analysis, resulting in a 12% attrition rate.

See also **Missing data**.

Morgan-Short, K., Heil, J., Botero-Moriarty, A., & Ebert, S. (2012). Allocation of attention to second language form and meaning: Issues of think-alouds and depth of processing. *Studies in Second Language Acquisition, 34*, 659–685.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*, 35–78.

Bb

β (see Beta)

Bar graph

A technique displaying quantitative data in which the scores of one or more variables are represented with columns. Generally, the y axis represents a variable expressed in counts (**frequency**) or as a **mean** score, and the x axis represents groups. Such displays of data are useful for providing a quick overall picture of the data. It is easy to see in the bar graph in Figure 1, for example, the relationship between participants' proficiency level and their scores on the two measures: the higher the proficiency, the higher the score, on average, for both the grammaticality judgment test (GJT) and for the error correction (EC) task, where learners are asked to identify grammatical errors in a written text. This bar graph also allows us to examine the relationship between participant scores and the choice of dependent measure. Regardless of proficiency level, the participants scored higher on the GJT than on the error correction task.

Despite their ease of use, bar graphs are not always an ideal way to present quantitative data. For example, bar graphs can be somewhat

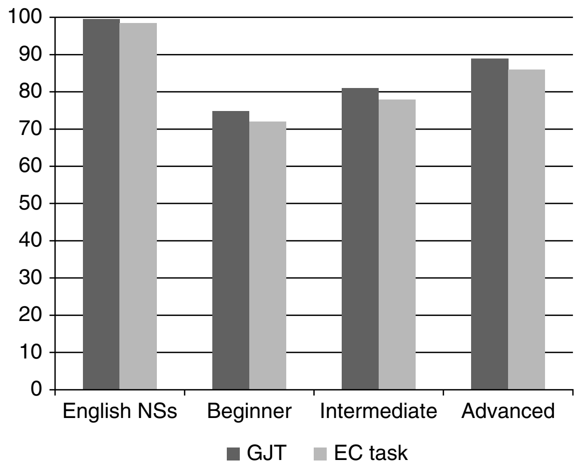

Figure 1 Bar graph

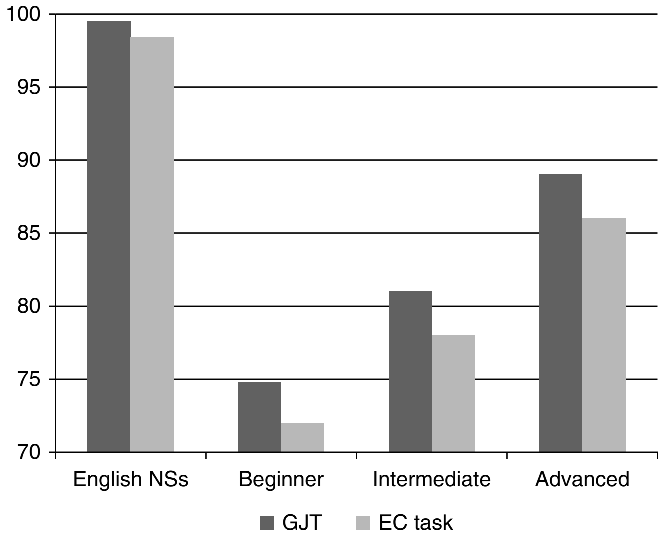

Figure 2 Bar graph with truncated scale

misleading if they do not include the entire measurement scale on the y axis. The abbreviated scale in Figure 2, where the scale begins at 70 rather than 0, presents an exaggerated view of the differences between mean scores compared with Figure 1, which presents the entire scale of 0–100. Another problem in most bar graphs, including the sample graphs here, is a lack of information about the dispersion in the data. By omitting such information, which can be included quite easily in the form of error bars, the reader is often led to overestimate group homogeneity (Weissgerber, Milic, Winham, & Garovic, 2015).

Weissgerber, T. L., Milic, N. M., Winham, S. K., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLoS One Biology, 13*, e1002128.

B

Bell curve

A term used in statistical analysis to refer to the shape of a **normal distribution** of scores when they are plotted using a **histogram**, **bar graph**, or **line graph**. A bell curve has its apex at the mean, which should be in the center of the graph, as shown in Figure 3. In addition, a bell curve has equal slopes descending from both sides of the apex, with the tails extending out on both ends. Visual inspection of data can provide information about their distribution, and if a bell curve is not present, then it is possible that the data are not normally distributed but positively or negatively **skewed.**