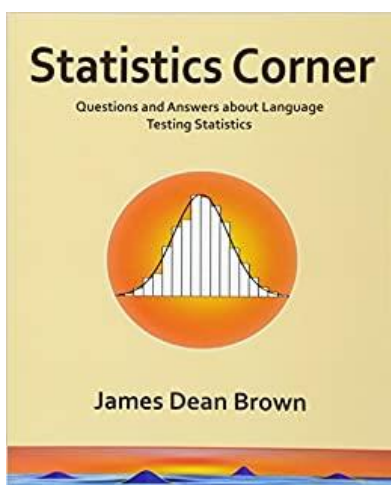




## Book Review

**Statistics Corner: Questions and Answers about Language Testing Statistics** by James D. Brown, JALT Testing and Evaluation Special Interest Group, 2016, 340 pp, \$28 (Paperback), ISBN-13 978-1537312866, ISBN-10: 1537312863 (eBook)



Statistics is concerned with collecting, organizing, summarizing, describing, and analyzing quantitative data for the purpose of interpretation and making decisions in almost all scientific fields. Statistical literacy is therefore a crucial component in and for academia, and language education is no exception. The applied linguistics community especially needs statistics for a) data analysis in conducting research and b) test score analysis in language testing. For both purposes, we need to make informed and objective decisions. Such decisions are based on evidence, which are provided by statistical analyses.

Statistical analyses have a key “role in the practice of language assessment” (Green, 2021, p. 25). Therefore, textbooks in language testing and assessment usually provide brief overviews of statistical techniques and their applications. Valuable volumes have also appeared with fuller treatments of the techniques useful for test validation (e.g., Bachman, 2004; Green, 2013).

*Statistics Corner: Questions and Answers about Language Testing Statistics* is an incrementally constructed repertoire of answers stemming over the years from the Statistics Corner, a regular column in Shiken: JALT Testing & Evaluation SIG Newsletter. The Newsletter was an early project of the Testing and Evaluation Special Interest Group of the Japan Association for Language Teaching (JALT), which aimed at encouraging the application of scientific approaches to language testing. Since the inception of the Statistic Corner, the author of the present volume, James Dean Brown, contributed voluntarily on a regular basis to answer questions about testing and statistics from interested readers, students, and editors. The questions came to Brown randomly and/or without any thematic order, and his answers were also largely informal. In 2016, Brown updated and arranged the content of the columns in a consolidated, thematic, and coherent book form, which allowed the content to spread to a wider audience.

Subsequently, the then 40-plus columns, except *eigenvalues* and *Yates' correction*, are converged fittingly into a book with two parts, each part with three sections that include chapters with cross-references. Table of contents, subject index, and other additions further the book's value as a handy, useful, user-friendly reference. Each chapter is structured in the following framework: title, questions raised, a succinct overview of the answer/solution(s) provided, an elaborated discussion of the answer/solution(s), and a conclusion. The chapters are at times an evaluation of a published study that poses a question to an interested reader, which then becomes the focus of a question-and-answer column of Shiken. Brown's responses could serve readers as practical guidelines in evaluating and/or conducting any study of the kind on their own.



The two parts of the book discuss second language testing and research. Section A in Part I, testing strategies, includes nine chapters. These chapters discuss 1) resources for interested readers, e.g., websites, journals, organizations, and books in language testing, 2) teachers' problems with classroom test writing, development, and validation, 3) differences in norm- and criterion-referenced test characteristics, development steps, and validation strategies, 4) construct validity (measurement), 5) two-stage testing, 6) intercultural pragmatics test development, 7) motivations of test-takers, 8) possible connections between extraneous variables and washback, and 9) strategies to promote positive washback impact from Japanese university entrance examination on English language education.

The chapters in Section B, Part I, touch upon item analyses, including 10) norm-referenced item analysis statistics (i.e., item facility and item discrimination), 11) criterion-referenced item analysis statistics (i.e., difference index and B-index), 12) use of point-biserial correlation coefficient in language tests, 13) distractor efficiency analysis using a spreadsheet to revise items, 14) ways of calculating statistics for items with weighted scores, 15) issues affecting Likert item questionnaire formats, and 16) confusion between Likert items and scales of measurement.

Section C of Part I (i.e., second language testing) is concerned with reliability issues. The chapters included relate to 17) appropriate use of Cronbach alpha, 18) optimum length of cloze tests, 19) difference between the standard error of estimate and the standard error of measurement, 20) use of Spearman-Brown prophecy formula for adjusting split-half reliability and addressing test length, 21) use of G-study and D-study for generalizability and decision studies, 22) process of calculating agreement and kappa coefficient for inter-coding, and 23) reliability of a teacher survey.

Part II is concerned with second language research and has three sections. In the three chapters of Section D, planning research, Brown defines qualitative, quantitative, and mixed-methods research as systematic and principled inquiries in terms of sound characteristics, or standards, that researchers and readers can use to defend and evaluate the research they conduct or encounter.

Section E of Part II, interpreting research, includes eight chapters (i.e., chapters 27–34). The questions addressed in these chapters are related to Chapter 27) six principles contributing to statistical research quality: distribution, assumptions, statistical significance versus meaningfulness, multiple statistical tests, causality, and null results, Chapter 28) how to generalize from second language research samples, Chapter 29) sample size and sufficiency of the power of statistical tests, Chapter 30) relation between sample size and statistical precision, Chapter 31) calculation of skewness and kurtosis, Chapter 32) importance of effect size and eta squared in applied linguistics, Chapter 33) difference between confidence intervals, confidence limits, and confidence levels, and Chapter 34) use of the Bonferroni adjustment to rule out spuriously significant differences in multiple comparisons.

Finally, Section F, research analyses, in Part II, has seven chapters that deal with Chapter 35) correlation coefficients versus coefficient of determination, Chapter 36) differences between principal components analysis and exploratory factor analysis, Chapter 37) different ways to decide the number of components or factors in principal components analysis or exploratory factor analysis, Chapter 38) types of rotation to choose in principal components analysis or exploratory factor analysis, Chapter 39) applications of principal components analysis and exploratory factor analysis in language-related research, Chapter 40) uses of principal components analysis and exploratory factor analysis in developing language tests and questionnaires, and Chapter 41) processes involved in using chi-square and related statistics for analyzing 2 x 2 contingency tables.

It is worthwhile to note that Brown had continued to contribute to the Statistical Corner column of *Shiken* since the current volume's publication in 2016. His last contribution appeared in *Shiken*, 23(2), December 2019. As such, future editions of this book should include these additions. It is regrettable that *Shiken* has stopped publishing the Statistics Corner column since then.

Due to the crucial role of statistical analyses in research studies and test scores, Brown has divided the content of the volume into two parts accordingly. As a result, the chapters do not seem to fit in the right part at first reading for the overlap between the application of questions and answers or use of analyses in both research studies and language testing. In spite of that, the running thread of the topic (i.e., statistical

analyses) keep the chapters together as a whole volume.

An update of the chapter contents would have improved the usefulness of the volume further. For instance, linking the questions and answers to the examples of studies recently published appear missing. The majority of the chapters in the volume looks similar to the way the contributions first appeared. The volume could include a new section about updates of the topic within the current framework of each chapter. Given the clear, simple, and informed style of Brown's writing, the new section would have added depth to the topics discussed.

The volume stands as testimony to how a needs-based approach can have an important role in addressing scientific issues among community members. In this case, the volume started from a column in a journal to serve the community's needs for statistical advice in 1997, accumulated the information to cover a wide range of statistical topics and address issues of concern among the community of language testing in this region and beyond, and finally consolidated in the form of a book. It is this book reviewer's view that the questions, though stemming from Japan, would be equally applicable to other EFL contexts, hence be of relevance to readers, teacher researchers, and graduate students around the world.

### References

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge University Press.  
Green, A. (2021). *Exploring language testing and assessment: Language in action* (2<sup>nd</sup> ed.). Routledge.  
Green, R. (2013). *Statistical analyses for language testers*. Palgrave Macmillan.

**Is'haaq Akbarian**

Department of English Language and Literature, University of Qom, Iran  
ishaaq.akbarian@gmail.com

(Received May 12, 2022; Revised July 10, 2022; Accepted August 14, 2022)