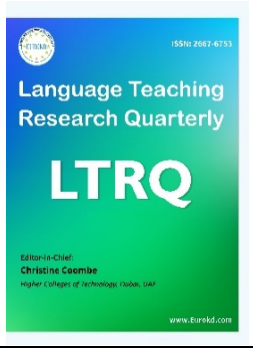


Language Teaching Research Quarterly

2023, Vol. 37, 4–75

James Dean Brown's 50 Years of Work in Second Language Studies: A Systematic Review

James Dean Brown[1], Ali Panahi[2], Hassan Mohebbi[3*]

[1]Professor Emeritus, University of Hawai'i at Mānoa, USA
[2]Iranian Foreign Languages Institute, Ardebil, Iran
[3]European Knowledge Development Institute, Turkey

Received *07 July 2023* ***Accepted*** *22 October 2023*

Abstract

Panahi and Mohebbi review James Dean Brown's 50-years of research in language testing, curriculum development and research statistics with reference to an impressionistic framework for analysis containing two components with their subcomponents: Annotations (i.e., briefing and implications) and main concepts and themes (i.e., testing and teaching terminology, research design, research instruments, data analysis, and domains). The review was carried out in two phases: In Phase I, we (Ali Panahi and Hassan Mohebbi) reviewed Brown's all works and extracted approximately 1100 main concepts and themes leading to 28 main entries for testing and teaching terminology. The issues he has examined are much more extensive; the first 10 topics and themes most widely investigated, in the order of frequency, are language testing and assessment, research and statistics, curriculum development and language programs, cloze tests, CRTs and NRTs, TESOL, ESL, applied linguistics and language testing, placement, standardized and proficiency tests, connected speech and reduced forms, pragmatics tests and issues, and reliability and validity. In Phase II, JD Brown provides a discussion and his personal reflections on the systematic review.

Keywords: *James Dean Brown, Systematic Review, Testing, Assessment, Research Statistics*

* Corresponding author.

E-mail address: hassan.mohebbi973@gmail.com

https://doi.org/10.32038/ltrq.2023.37.02

Introduction

The systematic review in Phase I goes beyond words of praise and reaches out to what Brown has done for the progress of language testing, research statistics, curriculum development and connected speech. All his professional works have significantly contributed to the field and brought about lasting changes to the field, especially by linking language testing findings to curriculum development and language pedagogy. Therefore, before presenting the framework for the systematic review, a brief account of some of his most important contributions are provided.

Stepping back historically, JD Brown's initiation into language testing profession (Brown & Hudson, 1998) was inspired by paradigm shifts in language testing ranging from discrete-point tests (the 1950s and 1960s) to the integrative tests (the 1970s and early 1980s), and then to the communicative tests (the 1980s and 1990s). As he points out (Brown, 2013a), his contributions started when he authored his master's thesis in the 1970s and was fascinated by cloze tests, leading to one apparent line of his work investigating cloze test issues. When in the 1960s and 1970s, there existed a discussion on the effectiveness of cloze as a test of overall ESL proficiency (Alderson, 1978; Oller, 1979), Brown reacted and argued that there was much variability in research findings associated with cloze tests reliability and validity; indicating that a black hole of information exists about cloze tests, he questioned the inconsistent results and announced a call for further research (Brown, 1984a). He has published a plethora of articles on cloze test, its development, use, scoring, reliability and validity. For example, *Relative merits of four methods for scoring cloze tests* (Brown, 1980) examines scoring methods in terms of reliability, validity, mean item facility and discrimination, or in usability associated with cloze test. Regarding the effectiveness of cloze tests, his article (Brown, 2002f) titled *Do cloze tests work? Or, is it just an illusion?* reveals the effectiveness of cloze tests in light of certain important variables. After 25 years of investigating cloze tests, he published a seminal article in 2013 titled *My twenty-five years of cloze testing research: So what?* (Brown, 2013a). He examined research works published between 1978 and 2002 on cloze testing and explored and reported the findings and eventually maintained that cloze tests function appropriately as one type of overall ESL proficiency tests.

From another relevant perspective, a historical review of his professional and personal growth elucidates that three academic events have coincided his professional development and ignited his interest in language testing and research statistics (Brown, 2017a): His first educational testing course with W. James Popham, i.e., mainly on criterion-referenced testing, his language testing and research statistics courses with Richard Shavelson, and through Shavelson, his introduction to the work of Lee J. Cronbach on generalizability theory. Affected by these scholarly experiences, he significantly contributed to the development of the statistical/quantitative terminology, research, and conceptual developments in applied linguistics. In short, we observed a motivating consistency in his attitudes towards research, language testing, professional development and publication trends, with which he moved an abundant number of researchers, testers, professionals, and educators. Curiously, after reading his earlier articles titled *Give second chances in education* (Brown, 2000f) and *Publishing without perishing* (Brown, 2005d) and his more recently published

articles titled *Professional reflection: Forty years in applied linguistics* (Brown, 2016a), and *Forty years of doing second language testing, curriculum, and research: So what?* (Brown, 2017a), and especially *JD Brown's essential bookshelf: Connected speech* (Brown, 2023), lead us to understand the scope of his professional contributions to the field and his personal growth. The fact is that reading his work, especially these five articles, clarifies the ins and outs of his academic efforts and reveals the fact that he openly paid tribute to those academic figures who had contributed to his own professional development in the field by crediting and citing their names in his articles—a strategy that would usefully be applied by other academics.

He has been constantly thinking reflectively about various gaps and concerns in the field leading to a set of contributions in their own right. For example, in one of his research projects titled *Resources on quantitative/statistical research for applied linguists* (Brown, 2004g), he notes that since his initiation into language teaching profession in the 1970s, he was extremely concerned about the poor quality of much of the quantitative and statistical research he read in ELT-related journals. Moreover, his book titled *Understanding research in second language learning: A teacher's guide to statistics and research design* (Brown, 1988c) and his informative articles such as *Statistics as a foreign language — Part 1: What to look for in reading statistical language studies* (Brown, 1991c) and *Statistics as a foreign language—Part 2: More things to consider in reading statistical language studies* (Brown, 1992h) were published to potentially fill this gap and serve the purpose, for both novice and more experienced scholars, of assisting them with using and interpreting statistical concepts associated with research design and data analysis. In particular, in 2013, he published an article for language teacher trainers about how to deal with teaching statistics in language testing course. In this regard, his article titled *Teaching statistics in language testing courses* (Brown, 2013i) is more informative.

On top of these all, an informative article titled *Designing a language study* (Brown, 1997a) appeared which covers some of the overarching concerns and issues in second language research. In addition, his practical experience in research has constantly led to useful insights for researchers. For example, in his article (Brown, 2012h) titled *What do distributions, assumptions, significance vs. meaningfulness, multiple statistical tests, causality, and null results have in common?*, he states his belief that after 35 years research, he had realized that scholars still do not understand the difference between significance and meaningfulness; they need to understand that the two are different things. These all are indicative of the fact that he did his academic share to improve the research competence of scholars in the field by struggling to add to and improve the statistical sophistication of both readers and researchers.

Throughout this review, we observed that Brown resisted drawing distinct borderlines between testing and assessment. For example, in their article titled *The alternatives in language assessment* (Brown, & Hudson, 1999), the authors clarify a number of misunderstood issues regarding testing and assessment. When Anthony Bruton criticized them (Bruton, 1999) maintaining that they were confused between testing and assessment, Brown and Hudson answered that it is not effective to draw an artificial borderline between tests and assessments because various forms of assessments

fall along a continuum ranging from discrete-point tests to more open-ended performance assessments and it is therefore in the responsibility of language teachers to make decisions as to what appropriate options to use in a specific situation for the purpose of language pedagogy in the classroom based on the particular needs of the learners.

One other thread of language testing research in the Brown's work was in the field of pragmatics; his article, titled *Pragmatics tests: Different purposes, different tests* (Brown, 2001a), provides a detailed and complementary account of the assessment of pragmatic proficiency already investigated by other researchers and examines and provides definitions for various pragmatics tests. Along this line, in relation to cross-cultural pragmatics, two books titled *A framework for testing cross-cultural pragmatics* (Hudson, Detmer, & Brown, 1992) and *Developing prototypic measures of cross-cultural pragmatics* (Hudson, Detmer, & Brown, 1995) tap into various issues such as the role of pragmatics in communicative competence, variables in speech act realization, instrument development process, and multiple methods for assessing cross-cultural pragmatic abilities.

In relation to university entrance examinations, he conducted local research in the Japanese context, for example, *English language entrance examinations at Japanese universities: What do we know about them?* (Brown & Yamashita, 1995a) and *University entrance examinations: Strategies for creating positive washback Strategies for creating positive washback on English language teaching in Japan* (Brown, 2000a). In accordance with his findings, the university entrance examinations can enhance positive washback effects on English language teaching, which has optimistic implications for instructors in terms of preparing much broader educational materials rather than narrowing down the content and limiting it to test taking strategies.

He has also made remarkable contributions to the field in terms of curriculum development and program evaluation. His first exposure to the notion of curriculum started when he was sent by UCLA to China to design, develop and implement an English for science and technology program in the Guangzhou English Language Center (Brown, 2017a). His seminal articles titled *Testing-context analysis: Assessment is just another part of language curriculum development* (Brown, 2008d) and *Forty years of doing second language testing, curriculum, and research: So what?* (Brown, 2017a) more clearly examine the close connection between language testing and curriculum design and their effectiveness, in particular the various stakeholders who are influenced by tests; that is why he has been an advocate of stakeholder-friendly curriculum, considering curriculum as an integral part of language assessment. On the nature of curriculum and language testing, he has widely presented, and published a plethora of journal articles, book chapters and books, to name just a few: *A systematically designed curriculum for introductory academic listening* (Brown, 1989f), *Placement of ESL students in writing-across-the-curriculum programs* (Brown, 1990a), *The social meaning in language curriculum, of language curriculum, and through language curriculum* (Brown, 1993a), *The elements of language curriculum: A systematic approach to program development* (Brown, 1995a), *The many facets of language curriculum development* (Brown, 2003g), *Language testing and curriculum development: Purposes, options,*

effects, and constraints as seen by teachers and administrators (Brown, 2002d), *Second language studies: Curriculum development* (Brown, 2006d), and many others.

Another main line of his work includes textbooks on language testing. A look back at the coursebooks available on language testing reveals that James D. Brown appears to be one of the main contributors whose assessment-related coursebooks are used for research, test development and use, and also for statistical research, such as *Testing in language programs* (Brown, 1996a). Additionally, he has created a plethora of new concepts and terminology, such as *testing-context analysis*, *stakeholder-friendly curriculum*, *stakeholder-friendly testing*, and *defensible needs-analysis-based curriculum* (Brown, 2008d), and the word *defend* (Brown, 2005b) in describing how validity arguments should be framed.

Brown's contribution is not solely limited to his publications; he has frequently presented at seminars and international conferences; his presentations on various language assessment and education issues published in conference proceedings are insightful and informative (e.g., Brown, 1983a, 1992e; Brown, Ramos, et al., 1991; Brown & Ross, 1993). Also, his joint work with another scholar entitled *The authors respond to O'Sullivan's letter to JALT Journal: Out of criticism comes knowledge* (Brown & Yamashita, 1995c) in response to O'Sullivan's letter shows how, while advocating for their research, they could maintain a healthy attitude toward criticism.

More remarkably, he can serve as a role model to researchers and professionals in the field in the way he admits to making mistakes in his insightful article titled *Language curriculum development: Mistakes were made, problems faced, and lessons learned* (Brown, 2009f). In addition, his recent article titled *JD Brown's Essential Bookshelf: Connected Speech* (Brown, 2023) demonstrates in action what he has expressed in words over the course of 50 years. Throughout his many articles, his passions for language testing, curriculum development, and research statistics (as well as many other related issues) are all clearly visible as shown in Tables 1-4 and Figures 1- 3.

To sum up, as the introduction above indicates, the issues he has examined over his career are extensive. To clearly present the topics and themes and to provide a comprehensive profile of Brown's work, the current study will cover the following topics: the organization of topics and research works, framework for the analysis, main concepts and themes, systematic review, concluding remarks for the systematic review, and references.

Organization of Topics and Research Works

We first created the needed inclusion and exclusion criteria and developed a framework for the analysis. In the process, it was necessary to clarify which research works to review, so we divided his research works for analysis into articles, book chapters, and books. Furthermore, considering the frequency, key roles, and pervasiveness of assessment, learning, teaching concepts, and research statistics notions in JD Brown's works, we created an impressionistic framework for analysis, so the selection was necessarily on a somewhat subjective basis. Admittedly, it is nearly impossible to specify the exact number of his works an author has published. Even Brown himself

was struggling to access some of his earlier publications, but due to the fact that they were much older and published decades ago, it would appear reasonable not to access all of them.

On these grounds, we reviewed 265 of his research works (Table 1) trusting that those reviewed would provide sufficient data for understanding his general contributions in some detail and therefore provide a clear representative profile of his contributions in terms of his total output and the topics he covered in his research. Moreover, owing to the circumstances of this study and space considerations, we removed from our systematic analysis (though we skimmed through all of it), his interviews, comments, research notes, brief reports, and summaries (with the exception of Brown, 1992f), responses to criticisms, the forums, more than 10 strings of book reviews, book notices, conference proceedings (with the exception of a few, such as Brown, 2003g), dissertations and theses supervised, video/audio presentations, and commentaries published on various aspects of qualitative and quantitative research, all of which would amount to a string of approximately 50 articles (e.g., Brown, 1983a,1983c,1990b, 1990e,1992b,1992e, 1992g, 1993b,1993c,1995b, 1995c, 1995d, 1995e, 1996b, 1996d, 2001b, 2002d, 2003d, 2003f, 2003h, 2005f, 2006f, 2006g, 2007c, 2021a, Brown, Romas et al. 1991; Brown & Ross, 1993; Brown & Salmani Nodoushan, 2015; Brown & Sato, 1990; Brown, Yamashiro et al., 2001; Brown & Yamashita, 1995c, 1995d; Lanteigne, et al., 2021b). Moreover, he has authored numerous book chapters, some of which appeared in the books he edited and published. So, we excluded from our systematic analysis a couple of book chapters including *Teacher resources for teaching connected speech* (Brown, 2006a), *Introducing connected speech* (Brown, 2006b) and *Testing reduced forms*, (Brown, 2006c) which were published in books edited by JD Brown himself, or jointly with others. Instead, we reviewed the main book in which those book chapters were published. Also, a closer review indicated that one research work was published as an article (Brown, & Hilferty, 1982), then, the same work appeared as a book chapter (Brown, & Hilferty, 2006); we reviewed the former and excluded from our systematic review the latter. Of course, it is important to note that we skimmed the entirety of his research work in order to provide a full picture of his contributions and the areas investigated.

As is clear in Table 1, the first 10 topics and themes most widely investigated with reference to his 265 articles, in the order of frequency, are language testing and assessment, research and statistics, curriculum development and language program, cloze test, CRTs and NRTs, TESOL, ESL, applied linguistics and language testing, placement, standardized and proficiency tests, connected speech and reduced forms, pragmatics tests and issues, and reliability and validity. See also Appendix 1 where we visualize the way we extracted the topics and themes for every individual research work. As is clear, there are 23 topics and themes in Table 1 and these various topics demonstrate JD Brown's wide range of interests. To present his contributions more visually, in order of frequency, we used a bar graph in Figure 1 for ease of comparison.

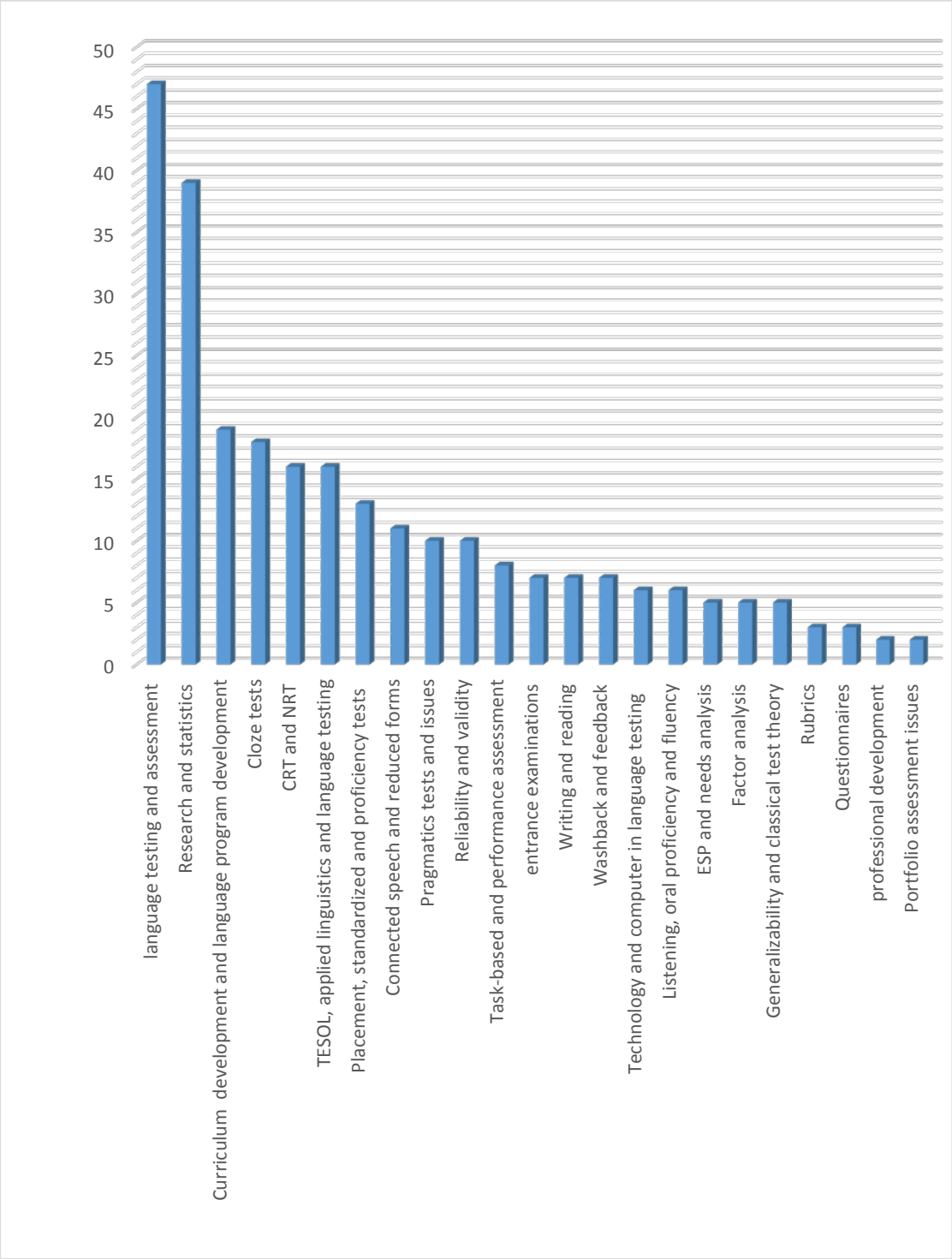
Table 1

Organization of the Vastness of his Contribution in the order of Dominance and Topic Types

Whole Number of Reviewed Articles: 265	
Primary Topics and Themes	Frequency
1. Language testing and assessment	47
2. Research and statistics	39
3. Curriculum development and language program development	19
4. Cloze tests	18
5. CRT and NRT	16
6. TESOL, applied linguistics, and language testing	16
7. Placement, standardized and proficiency tests	13
8. Connected speech and reduced forms	11
9. Pragmatics tests and issues	10
10. Reliability and validity	10
11. Task-based and performance assessment	8
12. Entrance examinations	7
13. Writing and reading	7
14. Washback and feedback	7
15. Technology and computer in language testing	6
16. Listening, oral proficiency, and fluency	6
17. ESP and needs analysis	5
18. Factor analysis	5
19. Generalizability and classical test theory	5
20. Rubrics	3
21. Questionnaires	3
22. Professional development	2
23. Portfolio assessment issues	2

Figure 1

JD Brown's Contribution in the Order of Dominance and Topic Types



A closer look at Table 1 and Figure 1 indicates that JD Brown's first five and most dominantly researched themes and topics are language testing and assessment, research and statistics, curriculum development and language program development, cloze tests, CRTs and NRTs. This is what he himself has mentioned in his various articles (e.g., Brown, 2009f, 2016a, 2017a) and the systematic review also provides a detailed account of these facts.

The Analysis

Establishing the impressionistic criteria for conducting the systematic review required subjective decisions on various issues. We adopted the "domain" section of the framework used by Fulcher et al. (2022) and added some further items to our domain list, i.e., papers on linguistic and sociolinguistic issues, papers on statistics, language education, and assessment research, and papers on professional reflection. The analysis contains eight columns: research works, briefing, implications, main theme, research design, instrument, data analysis concepts, and domain. Most of the articles were readably available and easy to download and obtain. However, as researchers have long found, one of the main problems with big contributors' research articles is that it is difficult, and in some cases impossible, to have easy access to their early work. We were in much closer online contact with JD Brown, and he provided us with those papers he could access. As a result, some of the articles were unavailable, even to the author, because they dated back four decades or more. Naturally, then, we necessarily excluded those articles which we could not access from our systematic review.

Google scholar and personal communication with JD Brown were the main sources for finding the works. Some of the articles had neither publication date nor the details and name of the journals in which they were published, so we excluded them from our analysis. Moreover, a few research reports of pilot projects were reviewed in the Book Analysis section (e.g., Brown, 2015d). As already noted, exclusion does not necessarily mean that we ignored the work totally, rather we skimmed those documents as well in order to confidently describe the landscape of his contributions to the field.

The review was carried out in two phases. In phase one, we (Ali Panahi and Hassan Mohebbi) systematically reviewed Brown's works spanning a period of almost 50 years including contributions to language assessment testing, curriculum development, and research statistics using an impressionistic framework for analysis containing two general components with their subcomponents: Annotations (i.e., briefing and implications) and main themes (i.e., testing and teaching terminology, research design, research instruments, data analysis, and domains). From his works, we extracted in total approximately 1100 technical concepts under the general heading of main themes leading in detail to their categorization into 28 main entries for testing and teaching terminology, all of which stood at 935 subentries (Figure 2), 4 general types of research design, 31 main entries for research instruments having 85 subentries, 18 entries for data analysis concepts with 80 subentries and also 17 domains. In phase II, Brown provides his personal reflections on this systematic review.

The analyses are presented also presented using tables and graphs created in Office Word 10. In Table 2 (*Analysis of the Articles*), Table 3 (*Analysis of the Book Chapters*), and Table 4 (*Analysis of the Books*), the publications are intentionally listed in chronological order to the extent possible. However, in a couple of cases, due to the existence overarching themes, the order was necessarily altered. Moreover, in most cases, we considered both main themes (i.e., research design, research instruments, data analysis, and domains) and annotations (i.e., briefings and implications), however in books and book chapters we only examined annotations. Before presenting the systematic analysis, we will describe the analytical framework in terms of main concepts and themes (i.e., research design, research instruments, data analysis, and domains).

Main Concepts and Themes

Main concepts and themes were prioritized for extraction on the basis of the frequency of the concepts and terminology as well as on an impressionistic basis. In deciding which research works to review, we divided his research works for analysis into articles, book chapters, and books. Of course, we skimmed and scanned the whole body of his research to present a due picture of his contributions. Considering the frequency, key role, and pervasiveness of assessment and learning and teaching concepts in JD Brown's work, we created an impressionistic framework for analysis, so the selection of the themes of the study was subjective. In what follows, we first present the main themes and provide a graphical analysis of the number of technical jargons in *Figure 2 and Figure 3.* Finally, we review and analyze the works, as is clear in Tables 2, 3 and 4.

Main Themes: Testing and Teaching Terminology and Concepts

1. ***Validity and test-use issues:*** validity, validity argument, content validity, construct and construct validity, systemic validity, face validity, criterion (or concurrent) validity, Messick's view of validity, divergent validity, convergent validity, discriminant validity, consequential validity, predictive validity, response validity (receptive-response types, productive-response types, personal-response type), internal validity, external validity, value implications, validation procedure, fairness and ethics, score meaning and inference, evidential basis, consequential basis, social consequences, practicality, usability, utility, interpretation, relevance, use, authenticity, evidence-based construct validity, validity statistics, construct generalizability, construct underrepresentation, content and truthfulness of test interpretation, credibility, confirmability, and transferability, replicability
2. ***Reliability and scoring issues*****:** reliability, rating scale or marking, analytic/holistic scoring system, holistic six point rating scale (0-5), 6-point rubric, holistic rubric, global (subjective) scoring, objective scoring, unitary rating, four to eight-point scale, primary trait scoring, T-unit concept for scoring, universe score, universe of observations, score descriptors or rubrics, oral interview (scale), score generalizability, cut score analysis, rater training, scoring rubrics, scoring cloze test, clozentopy scoring method, multiple choice scoring, exact answer, acceptable answer, test-retest, parallel forms, equivalent-form, split-half, internal consistency, readability scale, gain score, score reporting, dichotomously coded (right/wrong), inter-rater and intra rater reliability, agreement, Likert, test consistency estimates, cut-points, CUNY evaluation scale, CUNY reading assessment test, interval scale scores, combined reliabilities, cloze reliability, English proficiency rating, standardized scores, computer and scoring, self-ratings, rankings, and Q-sorts, Fry scale estimates, EFL difficulty estimate, exact-answer scoring method, score dependability, threshold loss agreement dependability, squared-error loss agreement dependability, domain-score dependability, classical test theory, grade point averages (GPA), consistency of measurement, separation reliability, rater agreement, coder agreement, stability of scores,
3. ***Test items:*** Item facility, item discrimination, item difference, item difficulty, item specifications, item types, item prototypes, item banking, item variety, discrete-point

items, integrative items, cloze items, nested items, fill-in items, translation items, passage-dependent item, passage-independent item, open-ended items, selected-response items, Likert-scale items, classical item analysis, reference items, substitution items, lexical cohesion items, conjunction items, non-technical vocabulary items, ten-item anchor cloze,

fact items, inference items, substantial vocabulary items, technical vocabulary items, scientific rhetorical function items, family of item types, item statistics, NRT item analysis, CRT item analysis, differential item functioning, local item dependence

4. ***Test types, formats, functions, related approaches, inventories and questionnaires:*** norm/criterion-referenced testing (NRT/CRT), domain-referenced testing, (modern) language aptitude test, communicative testing theory, two-stage testing, cloze test (word deletion pattern) tailored cloze, 7^{th} word deletion pattern, 12^{th} word deletion pattern, well-tailored cloze test, natural cloze test, sentential cloze test, inter- sentential cloze test, composition test, C-test, reduced-forms dictations, open-ended tests, constructed-response format, short answer test, performance tests, pragmatics tests, computer-based test, computer-assisted testing, computer-adaptive testing, multiple choice tests, flexi-level test, composition tests, large scale (high-stakes) tests, discrete point tests, placement test (placement test for reading/listening/writing/ speaking, writing-across-the-curriculum program) communicative oral test, computer-based test, web-based test, oral communication test, entrance examinations, receptive tests, productive tests, passage-based tests, integrative tests, true-false tests, achievement tests, dictation test, connected-speech narrative dictation test, connected-speech conversation dictation test, achievement test, diagnostic test, proficiency test, minimal competency test, limited English proficiency test, vocabulary test, role-play tests, task-based tests, diagnostic pretest, subtests, Yatabe-Guiford Personality Inventory, Attitude/Motivation Test Battery, Foreign Language Classroom Anxiety Scale, Strategy Inventory for Language Learning, Eysenck Personality Questionnaire, Bernreuter Personality Inventory, Guilford Personality Inventory for Factors STDCR, the Guilford and Martin Personality Inventory for Factors GAMIN, the Guilford and Martin Personnel Inventory, Portuguese versions of the Motivated Strategies for Learning Questionnaire, MSLQ, educational testing
5. ***Assessment, evaluation and decision issues:*** assessment, on-going assessment, alternative assessment, alternatives in language assessment, classroom assessment, classroom testing, performance assessment, formative assessment, formative evaluation, summative assessment, summative evaluation, assessment of oral proficiency interview or speaking, assessing writing, teacher evaluation, teacher assessment, selected-response assessments, true-false assessment, matching, multiple-choice assessment, constructed-response assessments, fill-in assessment, short-answer assessment, personal-response assessments, conference assessment, portfolio, portfolio assessment, self- or peer assessments, portfolio assessment, self-assessment, multiple-pairing strategy for writing assessment, (language) program evaluation, evaluation, process-based decision, product-based decision, field research-based decision, laboratory research-based decision, during-the-program evaluation, after-the-program evaluation, diaries, portfolio evaluation, teacher portfolio for evaluation, language

program decision, staff retention decision, cut-back decision, evaluation questionnaires, teacher evaluation, procedure, traditional assessments, triangulation of decision making process, language testing practices, assessment practices, assessment strategies, assessment of accuracy and fluency

6. ***Test construction issues:*** test purpose, test design, test delivery, test model, optimum test length, test specifications, test length, test development, test development practices, test writing practices, test validation practices, clear heading, clear directions, proofreading the test, numbers of items,
7. ***International tests:*** superordinate tests (TOEFL, TOEIC) standardized tests, Graduate Record Examination (GRE), IELTS (International English language testing system), Internet-based TOEFL, Hawaii Test of Essential Competencies, Test of English for International Communication (TOEIC), Interagency Language Roundtable Oral Interview, Oral Proficiency Interview (OPI), ACTFL, Educational Testing Service (ETS), TEEP, Hawaii State Test of Essential Competencies, Ontario Test of ESL, Common European Framework Of Reference (CEFR), AACES, COT, ELIPT, SPEAK, TOEFL, TOEIC, TSE, TWE
8. ***Standardization issues:*** standards, frameworks, standards setting, politics
9. ***Language education, curriculum and applied linguistics:*** language proficiency, English as a second language (ESL), English as a foreign language (EFL), informal speech, realistic pronunciation, contraction, assimilation, reduction, learning and teaching, input, output, fluency development, schemata, test taking strategies, test-taking abilities, teaching English to speakers of other languages (TESOL), TESL (teaching English as a second language), English language proficiency, continuing professional development, teaching and testing, English as a lingua franca (ELF), World Englishes, English as an International Language, inner circle of Englishes, outer circle of Englishes,, expanding circle of Englishes (Kachru 1986), global standard English, cheating, needs analysis, language needs, context needs, testing-context analysis, stakeholder-friendly curriculum, stakeholder-friendly testing, communicative and interactive strategies, type and token, learner autonomy, higher order cognitive skills, intensive reading strategy, extensive reading strategy, academic study skills, program level decision, classroom level decision, socioeconomic level, process-oriented approach to writing instruction, Bloom's (1956) taxonomy, higher-order skills, lower-order skills, students of limited English proficiency (SLEP), intensive ESL training, additional intensive ESL training, teacher training, computer assisted materials, media-centered materials, materials development, curriculum development, grammar-translation tradition, direct approach, audio-lingual method, stimulus-response notions, communicative syllabuses, task-based curriculum, technology, computer and language testing/teaching/learning, multimedia, supporting teachers, professional competence, grammatical competence, sociolinguistic competence, strategic competence, pragmatic competence, higher-order cognitive skills, syntactic complexity, T-unit, metacognitive, cognitive, and social-affective strategies, integrative motivation, instrumental orientation, parental encouragement, social competencies, Englishes in testing, grammar-translation method, communicative language teaching, functional syllabuses, skills acquisition, task-based activities,

TENOR (teaching English for no obvious reason), publishable papers (the field would read), term papers (professor would read), test-taker motivation, classical method, grammar-translation method, direct method, audiolingual method, cognitive code, communicative approach, structural syllabus, situational syllabus, topical syllabus, functional syllabus, notional syllabus, lexical syllabus, skills-based syllabus, task-based syllabus, teacher training, high proficiency students, low proficiency students, connected speech, publishing without perishing, pedagogy, political structures, discrepancy philosophy, democratic philosophy, analytic philosophy, diagnostic philosophy, purposes (CRTs/NRTs), Hawaiian language immersion program, cultural awareness, comprehensibility, intelligibility

10. ***Language skills and sub-skills:*** listening comprehension, reading, writing proficiency, speaking, pronunciation, grammar, vocabulary
11. ***Testing problems, constraints, sampling and various effects:*** rater effect, testing effect, main effect, practice effect, placebo effect, placebo, interaction effect, teacher-effect problems, test effect, method effect, Hawthorne effects, halo effect, novelty effect, sampling strategies, samples of convenience, random sample, sampling error, stratified random sampling, sample size, sample size effect, generalizability of the study, sampling and generalizability, teaching-style effect, teaching-strategy effect, counterbalancing effect, program-fair instrument problems, political-problems effect, subject expectancy effect, researcher expectancy effect, reactivity effect, mortality of participants effect, self-selection effect, Type I and Type II errors, strength of treatment, functional constraints, political constraints, economical constraints, undifferential error
12. ***Task issues:*** task type, task validity, anxiety (task anxiety, computer anxiety), language tasks, multiple tasks, task difficulty, essay type task, language type task, reading type task, task specifications, rephrase task, reorder task, short answer task, dictation task, authentic tasks, (analytical) writing task, open-ended narrative task, reading task, listening task, essay prompts and topics, open-response prompt, interview tasks, proficiency interview task, problem-solving tasks, communicative pair-work tasks, role playing tasks, group discussions tasks, real-life language tasks, task specifications, task content, task characteristics (setting, input, rubrics, expected response, input/response relation), task-dependent, task-independent, oral discourse completion tasks, self-assessment tasks, written discourse completion tasks, multiple-choice discourse completion tasks, discourse role-play tasks, self-assessment tasks, role-play self-assessments
13. ***Trait facets (ability continuum) method facets:*** test method, multi trait-multimethod matrix, hit or miss method, modification method, tailored cloze method, latent trait analysis, masters, non-masters, testing considerations, human considerations
14. ***Text issues:*** passage difficulty, text difficulty, (test) input, content words, function words, passage readability, authentic text
15. ***Validity threats: Systematic threat to fairness, threat to validity:*** norming bias, linguistic bias, cultural biases, construct-irrelevant variance, internal validity threats, history, maturation, testing, instrumentation, statistical regression, selection bias, experimental mortality, selection-maturation interaction, external validity threats,

reactive effects of testing, interaction of selection biases, treatment bias, reactive effects of experimental arrangements, multiple treatment interference, construct-underrepresentation, construct-irrelevant variables, environment of the test administration, administration procedures, examinees, scoring procedures, test construction, quality of test items

16. ***Testing-related models:*** participatory model, diagnostic feedback model, native speaker model
17. ***Specific purpose issues****:* English for specific purposes (ESP), English for academic purposes (EAP), subject matter, ethnicity
18. ***Language components:*** phonological, lexical, syntactic, semantic, pragmatics
19. ***Feedback and washback and strategie*s:** test feedback, washback, positive washback effect, negative washback effect, washback validity, backwash, test impact, test design strategies, test content strategies, logistical strategies, interpretation strategies, content analysis, classroom observation feedback, measurement-driven instruction, teaching to the test, prestige factor, accuracy factor, transparency factor, utility factor, monopoly factor, anxiety factor, practicality factor, curriculum factor, measurement facets
20. ***False beginners:*** The testing of false beginners
21. ***Examinations formats****:* oral, aural, written
22. ***Data issues in testing****:* cloze data, original cloze test data, tailored cloze data, piloted cloze data, pilot testing, nominal data, quantitative data, qualitative data, triangulation of data, performance-based data, sources of data, data from students, data from families, data from teachers, biodata survey, opinion survey, diaries, journals, logs, behavior observation, interactional analysis, inventories, participant observations, nonparticipant observations, classroom observations, in person data, telephone data, internet data, unstructured interview data, structured interview data, interview schedules data, data from Delphi technique, data from advisory meetings, data from focus group meetings, data from interest group meetings, data from review meetings, data from questionnaires, biodata surveys, opinion surveys, closed response, open response, closed-response self-ratings, open-response self-ratings, closed-response judgmental ratings, open-response judgmental ratings, closed-response Q sort, data from text analysis, data from discourse analysis, data from role plays, data from simulations, data from content analysis, data from register/rhetorical analysis, data from computer-aided corpus analysis, data from genre analysis, quantitative information, qualitative information, subjective information, objective information
23. ***Characteristics, variables and research types and design****:* test characteristics, linguistic variables, extralinguistic variables, dependent variables, independent variables, moderator variable, homogeneous variance, cognitive, affective, and personal variables, extraneous variables environment issues, grouping issues, people issues, measurement issues, experimental group, random experimental group, control group, random control group, true experimental design, post-test-only design, treatment issues, quasi-experimental group, quasi-experimental design, time series, pretest-posttest design without control group, action research, qualitative research, quantitative research, library research, laboratory research, document search, mixed-methods

research, meta-analysis, research synthesis, legitimation, ethnographic research, survey type research

24. ***Performance descriptors:*** cohesion, content, mechanics, organization, syntax, vocabulary
25. ***Statistical language, and concepts:*** experimental research, quasi-experimental research, statistical reasoning, interpreting statistics, descriptive statistics, exploratory statistics, exploratory factor analysis, inferential statistics, statistical differences, follow-up statistics, probability levels, hypothesis testing statistics, statistical tests, assumptions of statistical tests, degrees of freedom, *F*-statistic, *p*-value, statistical significance, meaningfulness, causality, multiple statistical tests, central tendency, structural equation modeling, dispersion, mean, standard deviation, significant differences, frequency, correlation coefficient, canonical correlation, random variation, chance factors, alpha level, chi-square test, *n*-way chi-square, McNemar test, Fisher's exact test, Pearson product-moment correlation coefficient or *r*, magnitude, ANOVA (one-way, *n*-way, repeated measures, *n*-way repeated measures, multivariate), Friedman one-way ANOVA, MANOVA, ANCOVA (multivariate, n-way, repeated measures, *n*-way repeated measures), *t*-test, simple regression, multiple regression, point-biserial correlation coefficients, loglinear analysis, logistic regression, phi coefficient, tetrachoric correlation, Cronbach α, Kuder-Richardson procedures, Kuder-Richardson 21 (K-R21), Kuder-Richardson 20 (K-R20), Cronbach alpha statistic, skewed distributions, Bartlett Box, homogeneity of variances, readability estimates, Gunning Fog readability, Flesch-Kincaid readability, the kurtosis statistic, leptokurtic, platykurtic, Spearman-Brown formula, split-half method, Guttman statistic, Fisher *z* transformation, statistical tables, attack strategies, variables of focus, dependent variables, independent variables, moderator variables, control variables, intervening variables, covariate, repeated covariate, independent covariate, scales, nominal scale (categorical and dichotomous scale), ordinal scale (continuous scale), interval scale, ranked and ratio scales, lick-it type scale, nominal variable, independent levels, repeated levels, comparison statistics (mean comparison, frequency comparison, and correlation coefficient comparison), multiple frequency analysis, Hotelling's *T*, Spearman rho, median test, *U* test, Kruskal-Wllis test, sign test, true dichotomy, Artificial dichotomy, linear, dimensional, factor analysis, multidimensional scaling, cluster analysis, one-way discriminant analysis, tried-and-true nonparametric statistics, Guttman scaling, path analysis, loglinear path analysis, independence of groups and observations, normality of the distributions, equal variances, linearity, non-multicollinearity, homoscedasticity, *F*-test, causal relationships, statistical abbreviation, column and row labels, abbreviations for variables, between-subjects comparisons, within-subjects comparisons, mean squares, sums of squares, degrees of freedom, residual, normal distribution, peaked distribution, percentile scores for initial screening, identification numbers, raw scores, percentile equivalents, histogram, item response theory (IRT), IRT analyses, The Flesch-Kincaid, Fog readability indexes, Flesch reading ease formula, Miller-Coleman Readability Scale, Fisher *z* transformation, *z*-value, first language readability indices, varimax rotation, principal component analysis, factor analysis of variables, factor loading, communities, Likert

items, scales of measurement, Likert-like item formats, two-stage Likert-like item formats, phrase completion Likert-like alternative, generalizability theory (G-study, G-theory, generalizability study; G-study), G-study stage, D-study stage, generalizability coefficient (G coefficient), dependability coefficient, GENOVA software, squared-error loss approaches, CRT difference index, *B* index, the posttest item facility, standard error of measurement, skewed distribution, distractor efficiency analysis, item analysis statistics, eigenvalue, two-factor analysis, three-factor analysis, point-biserial correlation coefficients, skew, kurtosis, skewness statistic, kurtosis statistic, skewed scores, standard errors of kurtosis, range, input range, output range, confidence level, confidence bounds, confidence and the *see*, confidence interval, confidence limits, multi-faceted Rasch model, FACETS analysis, multivariate and scalar analyses, coefficient of determination, Yates' correction factor, chi-squared analysis, power value, power statistic, statistical precision, parameters, Bonferroni adjustments, effect size, eta squared, partial eta, rotation, variance components, fixed facets, random facets, separation index, probability curves, Cohen's kappa, kappa coefficient, vertical ruler, probability curves

26. ***Associations:*** Language Laboratory Association of Japan, The Japan Association for Language Teaching, the American Psychological Association, the American Council on the Teaching of Foreign Languages (ACTFL)
27. ***Logistical issues:*** time, resources, economy
28. ***Individual differences:*** motivation, anxiety, personality, attitude, attribute, ability, skill, accountability, self-scrutiny

Research Methodology Concepts

Method design

1. Review paper, qualitative approach, descriptive type, and survey types
2. Quantitative approach
3. Mixed-method

Instruments:

1. Open-ended format
2. Multiple-choice format
3. (Experimental) cloze test, dictation
4. UCLA ESL Placement Examinations, Michigan Placement Test
5. Placebo lesson or task
6. Pre-test/post-test
7. Selected reduced forms
8. Discussion/argument-based vs. presentation-based
9. Integrative Grammar Test
10. UCLA English as a Second Language Placement Examination Listening Sub-Test
11. Academic Listening Test
12. Scoring based on organization, logical development of ideas, cohesion, content, organization, grammar (syntax), mechanics, and style and vocabulary.

13. Composition writing; essay writing, analytical writing task; analytic essay (reading-based/personal experience-based writing), writing sample, writing-testing, scoring materials for across-the-curriculum program
14. Open-ended comments, open-ended narrative task, speech acts, storytelling tests, Hawaiian Oral Language Assessment materials
15. Questionnaires (Likert-type scale, program evaluation questionnaire); holistic six-point rating scale, self-administered questionnaires, group-administered questionnaires forms,
    Personality Inventory, attitude/ motivation test battery, Oxford's Strategy Inventory for Language Learning, task-specific scales and criteria, holistic scales and criteria
16. Self-access reading materials,
17. (Michigan) placement test for reading, placement test for speaking, SEASSI placement test, SEASSI oral interview tests, placement test, interview, individual interview, group interview, telephone interviews
18. TOEFL
19. Fisher *z* transformation, readability scale,
20. Guidebooks, examination papers
21. The Flesch-Kincaid, Fog readability indexes
22. Review and reasoning based on testing theory and practice and research design and statistical language
23. Guangzhou English Language Center (GELC) Test
24. Engineering English Reading Test
25. Reading passages, reading comprehension test, word deletion patterns, biodata blanks, directions
26. Cloze tests, fifty cloze procedures, fifty randomly chosen books fifty passages, cloze passage
27. NRT and CRT item analysis approaches, NRT/CRT reliability
28. Machine-scorable answer sheet; test booklets
29. Sub-tests: listening comprehension subtest, structure and written expression subtest, vocabulary and reading comprehension subtest
30. Role plays
31. Self-reports, stimulated recall, verbalized strategy use, retrospective reports
32. Audio recordings, video recordings or CCTV

Data analysis

1. Descriptive statistics (frequency, average, percentage, mean, standard deviation, maximum score, minimum score, histogram, scatterplot); scalar statistics, inferential statistics; chi-square statistics; descriptive test statistics, cross-tabulation of features, descriptive testing characteristics
2. Reliability coefficients, classical theory reliability estimates
3. K-R20 formula, K-R21 formula
4. Spearman-Brown prophecy formula
5. Validity coefficients, threshold loss agreement approach, agreement coefficient, squared-error loss agreement approach, phi(lambda) dependability index, dependability approach, stepwise regression coefficients

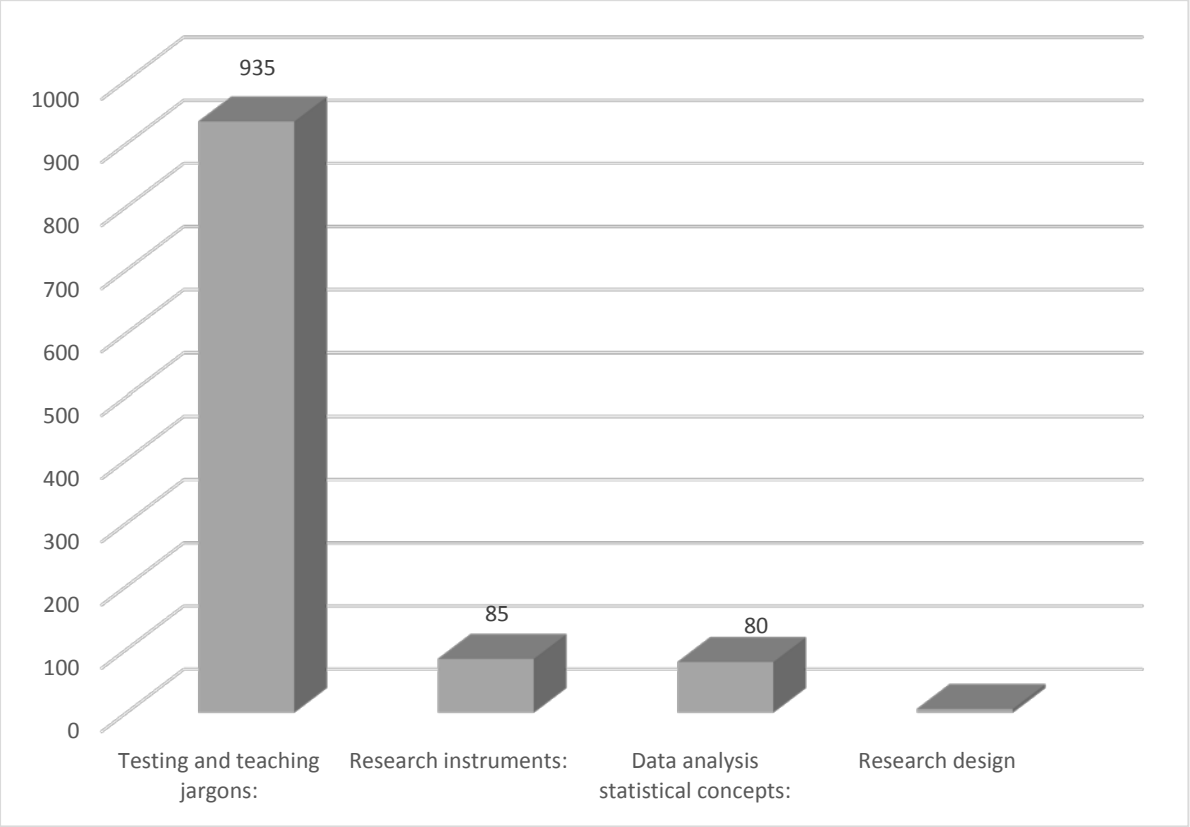
6. Correlation coefficients; correlation matrix, short-cut estimate, phi coefficient, Kappa coefficient, generalizability coefficient for absolute error
7. Open-ended scoring methods
8. Item analysis
9. Two-way repeated measures, three-way analysis of variance, and one-way ANOVA, multivariate analyses, MANOVA, SPSS, GENOVA, an overall repeated-measures analysis of variance (ANOVR), Wilks' lambda, Hotelling-Lawley trace *F* statistics, univariate statistics, Scheffe post hoc comparison, multi-faceted Rasch model, multivariate and scalar analyses,
10. Generalizability theory, generalizability studies, FACETS analysis
11. (Multiple) Regression analysis, SPSS plot, SPSS regression, varimax rotation, principal component analysis, factor analysis of variables, factor loading,
12. Rater questionnaire analysis
13. Computer analysis: the Quattro spreadsheet program, ABSTAT statistical program, SYSTAT statistical analysis programs
14. Right Writer computer program
15. Evidential, and review-based analysis or statistical reasoning
16. Cronbach alpha, split-half method, Pearson product-moment correlation coefficient
17. Item discrimination
18. Fisher’s *t*-test, *t*-test, multiple *t*-test; Type 1 error, independent means, nondependent means; *F*-test

Domain

1. Papers on validity, reliability, rating scales, scoring and performance tests
2. Papers on language testing, language education and technology
3. Papers on test design and development and test types
4. Papers on language testing and assessment

Figure 2

The Total Number of Technical Jargons



As Figure 2 indicates, 935 technical jargons for testing/teaching, 85 jargons and concepts for research instruments and 80 items for statistical concepts were extracted; the total number of the terminology for the main themes and the subcomponents stood approximately at 1100 technical jargons.

Figure 3

The Number of Technical Jargons and Concepts based on the Main Themes

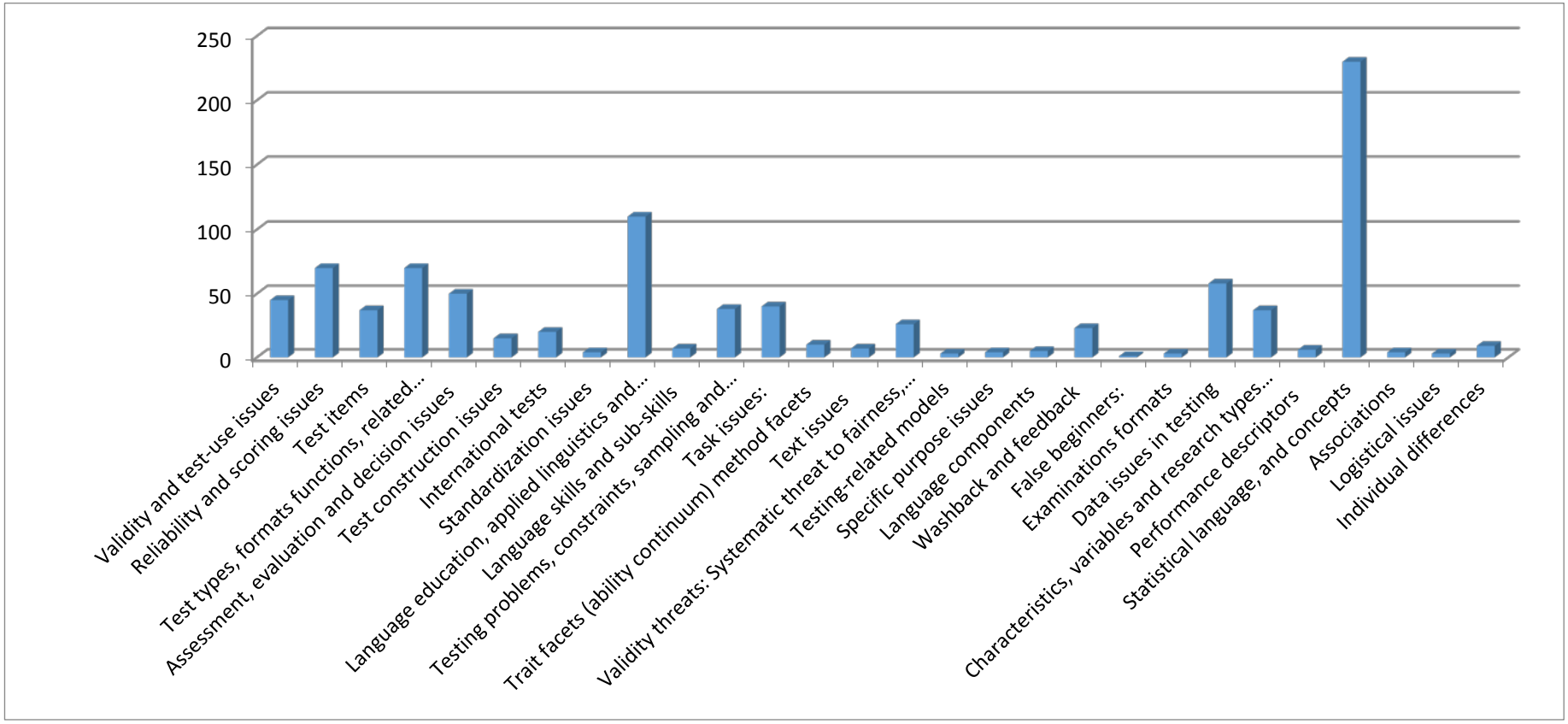


Figure 3 displays 28 subentries and the issues investigated. The individual subcomponents of the main themes considered, some of the issues were highly researched. For example, much larger number of statistical languages, concepts and jargons and language education and applied linguistics items were observed in his research works.

Table 2

Analysis of Articles

Annotations			Main Concepts and Themes				
Articles	**Briefing**	**Implications**	**Testing/ Teaching Terminology**	**Research design**	**Instruments**	**Data Analysis**	**Domain**
Brown (1980)	This study compared four scoring methods in terms of reliability, validity, mean item facility and discrimination, and usability: The results revealed and discussed differences among the four scoring methods.	Test designers can use the four methods with reference to the needs and purposes of the target population for whom the test is designed.	1, 2, 4, 9, 25	1, 2	1, 2, 3, 4	1, 2, 3, 4, 5, 6, 7, 8	1, 3, 4
Brown & Hilferty (1982)	This classroom research paper investigates the effectiveness of teaching reduced forms for developing listening comprehension: Results indicated that it is effective.	Language teachers can teach reduced forms and testers also can do further research on measuring other reduced forms.	7, 9, 10, 11, 25	2, 3	5, 6, 7, 8, 9, 10	1, 9	4, 9, 14,15
Brown & Bailey (1984)	This study examines a categorical instrument for evaluating compositions with five benchmarks for scoring.	Teachers can use this scoring instrument on compositions for measuring ESL learners' writing proficiency.	1, 2, 4, 8, 10, 16, 25	1, 2, 3	11, 12, 13, 14	2, 10, 11, 12	1, 4, 7,
Brown, Yongpei, & Yinglong (1984)	The influence of self-accessed reading materials in an English for science program was evaluated; self-access materials turned out to be of actual and potential benefit.	Teachers can provide learners with self-access materials and evaluate those self-access materials.	1, 2, 4, 9, 10, 17, 25	2, 3	2, 14, 15, 16, 17, 18	3, 9	1, 2, 4, 6, 10
Brown (1988a); Brown & Grüter (2022)	The likelihood of improving the reliability and validity of a cloze procedure was investigated with the use of traditional item analysis and selection techniques. In this regard, Brown and Grüter's (2022) more recent work on cloze test issues is more insightful.	Future scholars can investigate the extent to which individual item facilities and discrimination values change as a function of the deletion surrounding the items.	1, 2, 3, 4, 7, 9, 10, 13, 14, 22, 25	2	3, 22	1, 2, 3, 4, 5, 16, 17, 18	1, 5, 9, 10
Brown (1988b)	This study investigates the variance components of three engineering-reading tests made up of 60 multiple-choice items. The results indicate that the engineering reading test is dependable and valid for measuring engineering English reading.	For measuring and boosting the engineering reading proficiency of learners, teachers can provide learners with pre-requisite general English materials, too.	1, 2, 3, 4, 10, 14, 17, 25	2	3, 23	1, 3, 6, 16	1, 4, 6, 10

Brown (1989b, 1988d)	The link between item difficulty and the linguistic characteristics of cloze test items was examined. The subjects were limited to only Japanese students, so generalization of results should be taken cautiously.	Researchers can replicate the studies in other situations with much larger sample sizes, and with other linguistic and extralinguistic variables.	1, 2, 3, 4, 14, 23, 25	2	18, 24, 25	1, 11, 16	1,4, 10
Brown (1989c)	This article investigates and compares the performance of native speakers and international students at the end of training in their respective composition courses: No significant difference was observed in the performance of the two groups.	Implications of the study can help language testers to understand writing features and performance descriptors for assessing the writing of both groups.	2, 8, 9, 10, 11, 12, 23, 24, 25	2	11, 12, 13, 14, 24	1, 3, 9, 16	1, 5, 15
Brown (1989d)	This study develops a placement test for ESL reading curriculum at the University of Hawai‘i and succeeds at developing and presenting a practical and useful model for developing program-related placement tests. Further related issues are also detailed in Brown, Hudson, et al. (2004).	Teachers and test designers and developers can use this model and the way it was created and to develop other placement tests for other language programs.	1, 2, 3, 4, 7, 9, 10, 25	2	3, 6, 10, 12, 16, 17, 24, 26	1	10
Brown (1989e)	This study reviews and surveys the reliability of criterion-referenced tests and their reliability estimates, and characteristics, as well as the importance of norm-referenced and domain-referenced tests (Brown, 1984b).	The study can help the testers and researchers to be aware of the ins and outs of dependability estimation and phi-coefficients.	1, 4, 25	1, 2	21, 26	2, 3, 5, 15	1, 3, 4, 16
Brown (1990a)	This study examines a writing-across-the-curriculum program associated with two placement tests (MWPE and ELIPT): Learners take one of six composition courses and develop the quality of their writing.	Teachers can apply a writing-across-the-curriculum approach to making placement decisions, as the related placement tests proved accurate and effective.	1, 2, 4, 9, 12, 25	2	11, 12, 14	1, 2, 3, 9, 13	1
Brown (1990c, 2003d)	This study reviews the significance, use, and characteristics of criterion-referenced tests in East Asian EFL contexts (Brown, 2003d) and describes four easy-to-calculate techniques for estimating the consistency of such tests.	CRTS, due to their direct relationship to learning, can be of potential benefit for investigation, and development of curriculum in language classrooms.	2, 4, 11, 13, 16, 25	1, 2	21, 26	5, 6, 15	1, 3, 4
Brown (1990d)	This study reviews and surveys the four types of decision-making processes (proficiency, placement, diagnostic, and	Teachers can learn to use tests as tools for supporting both teaching and learning and making	4, 7, 9, 12, 13	1	8, 21, 26	15	1, 4, 7, 9

	achievement) used in any language teaching institution; each is described in terms of NRT and CRT.	classroom level or program level decisions.					
Brown (1991a)	This article investigates the Hawaii State Test of Essential Competencies as a minimal competency test; students must pass this test to graduate from high school.	Teachers and administrators must consider students' backgrounds in terms of language, culture and ethnicity in competency testing decisions.	1, 2, 3, 4, 7, 17, 19, 25	2	27	1, 9, 13, 18	1, 2, 3, 4, 15
Brown (1991b)	This article investigates whether and how two populations of students (ESL vs. native speakers) differed in their writing performances at the end of freshman composition training: The differences were observed.	Workshops and cooperative actions can be taken in order to teach the raters as to how to rate various essays and this can lead to more accurate scoring.	2, 9, 10, 23, 24, 25	2	11, 12, 13	1, 3, 9	1, 4, 7, 11, 12
Brown (1991c)	This study examines statistical reasoning, concepts and issues and provides EFL/ESL teachers and researchers with statistics-related attack strategies for reading, analyzing, and understanding research papers in the field.	For successful language education, teachers and novice scholars need to understand research design and statistical concepts to maximize their effectiveness with EFL and ESL students.	9, 18, 24, 25	1	21	15	16
Brown, Hilgers, & Marsella (1991)	This study investigated the degree to which individual prompts and topic types influence performance on the Manoa Writing Placement Examination. Brown (1989h) is informative on these same issues.	Teachers can assess learner's writing performance through their portfolios, and on the basis of a multiple-pairing strategy for prompts.	2, 4, 5, 12, 15, 24, 25	2	12	1, 9	1, 4, 7
Brown (1992a)	This study explores issues related to the role of computers in language testing. It deals with issues such as scoring, managing the results, placement testing, reporting the results, item banking, test administration, computer adaptive testing, and the various advantages and disadvantages.	Language teachers and test developers should consider the significant advantages of computers as useful tools, but subservient to the instructors needs and to the curriculum elements and assessment goals.	4, 7, 9, 11, 12, 13	1	21	15	2
Brown (1992c)	This study reviews various issues on language education/assessment and technology in two arenas (the media, and the method), and draws conclusions. More recently, Brown has considered the	Teachers, test developers, and administrators need to be technologically literate in order to appropriately evaluate programs	4, 5, 6, 9, 10, 26	1	21	15	2,4

	stakeholders' views views of technology and its contributions in language learning (Trace & Brown et al., 2017a, 2017b).	and assess learning outcomes appropriately.					
Brown (1992d)	A questionnaire-based study conducted by the TESOL Research Task Force surveys reviews issues related to the definition of research.	The answers obtained have implications for teachers, as they can inspire teachers to do action research.	23	1, 2	21	15	1, 4
Brown (1992f)	This study examines which characteristics of students of limited English proficiency (and performance on the Hawaii Test of Essential Competencies) can predict performance on other standardized tests.	Teachers can use the content of minimal competency test to help their learners meet their minimum communicative needs.	2, 4, 7, 9, 25	2	21	1	1, 4
Brown (1992h)	This study examines and reviews a complex subject area, i.e., statistical concepts and statistical language research conducted in the context of those practicing and researching in the field of EFL/ESL.	Researchers and teachers can use more advanced strategies required for reading, understanding, and analyzing research articles in statistical terms.	4, 9, 11, 25	1, 2	21	15	4, 16
Brown (1993d)	This study examines the characteristics of natural cloze tests with use of fifty randomly selected reading passages and a sample size of 2298 EFL students in Japan. The results indicate that natural cloze tests are not necessarily well-centered, reliable, and valid.	Teachers or test designers should not simply select a passage and develop a cloze test from it. For a cloze test to function, it should be tailored.	1, 2, 3, 4, 7, 25	2	24	1, 9	1, 16
Brown (1995f)	This study discusses language program evaluation, including issues like summative/formative decisions, participatory model, field research, or laboratory research, quantitative/ qualitative, or product/process data, quantitative/qualitative data for evaluation, sampling, the practice effect, and the Hawthorne effect.	Teachers can use the evaluation issues in the study and consider language or program evaluation a practical activity. They can then use need-based data to bring about effective change to language teaching and language curriculum.	4, 5, 9, 10, 11, 16, 22, 25	1	21	15	4, 5
Brown (1995h)	This study reviews special exam- related vocabulary in Japan because teachers need to understand such terms in order to understand the entrance examination system that affects many of their students.	The study has important implications in the Japanese context. Teachers need to perform various activities in the class in	4, 9, 10,	1	21	15	4, 11

		order to help their students cope with entrance examinations.					
Brown (1995i)	This satirical study examines the end-all TESOL Questionnaire for Investigating Really Kaleidoscopic Samples (aka, QUIRKS [pronounced quirks] and makes fun of statistical studies in general.	The study will be funny to the degree that the reader understands research methodology and statistics.	23, 25	2	14	1, 2, 16	1, 4
Brown & Yamashita (1995a)	This article investigates ten examinations each from private and public Japanese universities item-by-item and examines the difficulty of reading passages and the items associated with the examinations.	The article shows some of the oddities of the English exams in Japan and argues that English language teachers should not limit themselves to teaching to the entrance examinations.	3, 4, 6, 9, 10, 12, 13, 18	2	19, 20	13, 14	1, 3, 4, 10
Brown (1996c)	This study reviews key issues on fluency development and examines issues such as linguistic prerequisites for fluency development, learning to make errors, and generating opportunities, as well as activities for fluency development.	Knowing that fluency is acquirable can help teachers to create speaking classes and learn how to control the class while minimizing teacher talk and maximizing learner talking time.	5, 9, 10	1	21	15	4, 8
Brown, Robson, et al. (1996)	This article examines personality, motivation, anxiety, strategies, and multiple measures of language proficiency all at the same time in Japanese context. The results indicate that the measures turned out to be highly effective and reliable.	Teachers can use the findings to help them understand students' individual differences. Also, they can recognize the traits of a good language learner.	4, 9, 25, 28	1, 2, 3	4, 14	1, 9, 11, 16	1, 4, 16
Oliveira et al. (1997)	This study surveys key language testing concepts and examines the effectiveness of the English Language Placement Test administered at the Catholic University of Rio de Janeiro to computer science students.	Test developers need to consider the use and efficacy of item analysis and test revision techniques for improving tests.	1, 2, 3, 4, 9, 25	2	16	1, 5, 8, 17	1, 3, 4, 6
Brown (1997a)	This article reviews some of the crucial issues in second language research by covering wide-ranging topics such as sampling, types of variables, research design, validity in research, and ethics.	The article can be effective for university professors, instructors, novice and experienced scholars in helping them conduct research on language teaching and learning.	1, 7, 9, 11, 15, 23	1, 2, 3	21	15	5, 16
Brown (1997b)	This study examines the reliability of surveys with reference to the significance and variation in the idea of consistency by	Developing any questionnaire requires careful examination of reliability because inconsistency	2, 4, 5, 7, 25	1	21	15	1, 16

	addressing questions concerning internal consistency estimates and the adequacy of a reliability value of 0.70.	in any instrument call the results into question.					
Brown (1997c)	This study examines different types of surveys including both interviews and questionnaires, as well as their functions, roles, and effectiveness in language curriculum development and research.	Teachers can use the informative content of this article to learn how to gather information for decision making in language classrooms and programs.	2, 5, 9, 22	1	14, 16	15	4
Brown (1997d)	This study reviews issues on washback in language education and covers important topics like the effectiveness of washback and factors affecting washback.	Teachers and educators need to consider the impact of testing effects and washback on language learning and teaching.	1, 2, 4, 7, 9, 10, 14, 15, 18, 19	1	21	15	1, 4
Brown (1997e)	This study reviews issues of skewness and kurtosis and supplies some essential rules and guidelines for using and interpreting the skewness and kurtosis statistics.	The study has implications for researchers to help them analyze and interpret their skewness and kurtosis statistics and reflect on their data distributions.	4, 25	1	21	15	1, 16
Brown (1997f, 1997g)	This article examines the nature, and effects of washback, factors affecting the impact of washback, negative aspects of washback, and ways to promote positive washback.	The study provides a solid background on the ins and outs of washback for researchers, teachers, and scholars.	9, 19	1	21	15	4
Brown (1997h)	This study reviews recent developments in the use of computers in language testing, focusing on item banking, computer-assisted language testing, computer-adaptive language testing, and research on the effectiveness of computers in language testing.	The study has implications for teachers, educators and researchers, in that it can create a basis for using computers in language testing, pedagogy, and research.	2, 3, 9, 11, 25	1	21	15	1, 2, 4
Brown & Wolfe-Quintero (1997)	This study discusses teacher evaluation processes including portfolio evaluation, resume content, teaching philosophy, conference presentations, and other practice-based activities.	Educators can use the content of this article to improve how they present a professional image, which can in turn contribute to their overall professional development.	5, 9, 19	1	21	15	4

Brown (1998b)	This study examines three factors influencing cloze tests reliability: changes in numbers of items, variations in student ability levels and score ranges, and differences in passage difficulties.	Test developers and language teachers can use the study to develop effective cloze tests.	2, 3, 4, 25	1	21	15	1, 3, 4
Brown (1998c)	Readability and its relationship to EFL students' performance on cloze passages was explored. Specifically, it examines the relationship between first language readability estimates and actual performance-based passage difficulties.	One of the implications is that the results can be used for estimating the readability levels of reading passages for EFL/ESL students.	2, 4, 9, 10, 25,	1	18, 25	1, 4, 8, 9, 11,	1, 4, 10
Brown & Hudson (1998)	This study presents the advantages and disadvantages for different types of language assessments and tests that teachers need to use; it also elaborates on the effectiveness of feedback and multiple sources of information for decision-making.	Teachers should consider all test types as alternatives in assessment and their usefulness depending on the purpose of the course, as well as the needs and interests of the learners.	1, 2, 4, 5, 8, 9, 10, 12, 13, 15, 19, 22, 27	1	21	15	1, 3, 4
Wolfe-Quintero & Brown (1998)	This study reviews issues on the nature of portfolios, teacher portfolios, portfolio content, sample items for a teacher portfolio, the uses of portfolios, professional development, student mentoring, and teacher evaluation.	Teachers and teacher trainers can learn the value of teacher portfolios and how to create and use them effectively.	5, 9	1	21	15	4
Brown (1999a)	Using a large sample size of 15000 test takers, this study examines the relative significance of TOEFL score dependability and the relative importance of items, subtests, persons, and languages, and their interactions.	The findings of the study can be useful for the improving the design and development of computerized and paper-and-pencil versions of the TOEFL.	3, 4, 7, 10, 18, 25	2	28	1, 2, 10	1, 3, 4
Brown (1999b)	This study reviews the effects, purposes, roles, and responsibilities in language testing by examining test types, validity factors, language test use and interpretations, and various perspectives on test validity.	Teachers can use the findings of the study in developing a test or in bringing about reform and changes to language testing and teaching.	1, 2, 4, 5, 9, 7, 19	1	11	15	1, 4, 5
Brown (1999c)	This study explains the standard deviation, standard error of the mean, standard error of measurement, and standard error of estimate and unscrambles the confusion often	The study has implications for researchers and language testers, as the study provides them with detailed explanations of the	25	1	21	15	16

	observed between standard error of estimate and standard error of measurement step by step.	differences among these various statistics and how they should be interpreted.					
Brown, Yamashiro & Ogane (1999)	This study examines the impact of the hit-and-miss method, modification method, and tailored-cloze methods for boosting the efficiency and effectiveness of cloze tests.	The results can help researchers (more experienced and novice ones) choose among cloze test development strategies in various settings.	1, 2, 3, 4, 7, 13, 25	2	25, 26	1	1, 3, 4
Brown (2000a)	This article examines the ways the university entrance examinations in Japan can enhance positive washback effects on English language teaching.	The central message: test designers and instructors should collaborate and inform each other of positive washback effects.	1, 2, 3, 4,5, 9, 10, 12, 19, 20, 21	1	21	15	1, 3, 4
Brown (2000b)	This study examines the general type of questionnaire item called a Likert-scale item and the factors which influence Likert-scale formats.	Researchers and teachers can use the article help them in designing and developing Likert-item questionnaires.	3, 7, 25	1	21	15	1, 16
Brown (2000c)	This study examines the concept of validity, its definition and provides an account of various types of validity with a focus on Messick's idea of validity.	Test developers should consider consequential validity, test use and interpretation and value implications in test design.	1, 2, 7, 13, 15, 22, 25, 28	1	21	15	1, 3, 4
Brown (2000e)	This study examines and reviews a coefficient alpha reliability estimate, which is one of the most commonly reported reliability coefficients, as well as the kinds of test it should be applied to and how it should be interpreted.	Researchers can use the findings of the study to help them interpret Cronbach alpha reliability coefficients in assessing the consistency of a set of items.	2, 25	1	21	15	1, 16
Brown (2001c)	This study provides a definition for point-biserial correlation coefficient, its relationship with other correlation coefficients, the calculation of the point-biserial correlation coefficient and its uses in language testing.	Language testers can learn how to assess the degree of relationship between a naturally occurring nominal scale and an interval or ratio scale.	3, 4, 25	1	21	15	16
Brown (2001d)	This study examines the nature Spearman-Brown prophecy formula and its use for adjusting split-half reliability, and for answering what-if questions concerning test length, test design, and test revision.	Test designers and teachers (interested in statistics) can use the findings of the study for revising and designing their own tests.	2, 4, 7, 25	1	21	15	16

Brown (2001f)	This study examines eigenvalues which are reported in factor analysis. Also, the study elaborates on factor analysis in language testing.	Testers and researchers can benefit from the brief overview of eigenvalues and factor analysis provided by this study.	25	1	21	15	16
Brown (2001h)	This study examines issues related to two-stage testing in norm-referenced testing, proficiency tests, placement tests and computer adaptive testing.	Teachers can use two-stage testing for norm-referenced purposes such as proficiency and placement testing.	4, 6, 25	1	21	15	1, 2, 3, 4,
Brown (2002b)	This study examines the topics of washback effect, negative and positive washback, test impact, measurement-driven instruction, extraneous variables, and curriculum related issues.	The study can raise teachers' awareness of the impact of tests on learning and teaching and of the areas needing remediation and progress.	1, 11, 15, 19, 23, 25	1	21	15	4, 16
Brown (2002c)	This study examines distractor efficiency analysis with examples, calculation of the item analysis statistics, and provides information to help in deciding which options and items are effective and which are not. Brown (2002e) is more informative on the Cronbach alpha reliability estimate.	Teachers and test designers can use distractor efficiency analysis as a useful tool for spotting miskeyed items and for tuning up ineffective items.	3, 4, 25,	1	21	15	16
Brown (2002f)	This article combines and reanalyzes the data from Brown, Yamashiro, and Ogane (1999, 2001) and investigates what it is that helps items function well in a cloze test at varyious proficiency levels.	The implications are that despite having negative aspects, cloze tests can potentially be characterized as just another test development technique.	2, 3, 4, 9, 25	1, 2	21, 24, 25	1, 3, 8, 9	1, 4, 16
Norris, Brown, et al. (2002a, 2002b)	This study reports on an investigation into the use and development of a prototype English language task-based performance test, especially the relationship between estimates of task difficulty and the performances of examinees.	Teachers can use the implications for assessing the task-based performances of the test takers based on more than one single rating scale.	2, 4, 5, 9, 12, 13, 22, 23, 25	2	14	1, 6, 9	1, 4, 16
Brown (2003a)	This study examines item analysis of norm-referenced tests (NRTs), i.e., item facility and item discrimination and covers various issues such as the overall purpose of item analysis, and item analysis statistics for NRTs.	Test developers can use these statistics for developing and analyzing norm-referenced tests, such as proficiency tests (e.g., IELTS, and TOEFL iBT).	3, 4, 25	1	21	15	16

Brown (2003b)	This study examines the distinction between a difference index and B-index and indicates that the two are used for analyzing CRT items for the purpose of revising the test. For producing curriculum and CRTs that match each other, both indexes are useful. Further accounts of the issue are informative in Brown (1991d).	Researchers or teachers can use the difference index to assure that items reflect the materials and the B-index for decision making at a certain cut-point.	3, 4, 25	1	21	15	1, 3, 4, 16
Brown (2003c)	This study examines and clarifies the coefficients of determination for cloze tests and explains coefficients of determination and the way they are calculated.	Researchers and language testers can use the article for calculating coefficients of determination.	4, 7, 25	1, 2	21	15	1,16
Brown (2003g)	Various elements of curriculum development are reviewed in terms of language tests, course and program development, the five historical approaches, basic philosophies of curriculum development, and data collection tools.	Implications apply to administrators, teachers, and researchers interested in the ins and outs of numerous aspects of curriculum development.	2, 9, 22	1	21	15	4
Brown (2003i)	This study reviews the significance of CRTs/NRTs and reviews related tests such as aptitude tests, proficiency tests, placement tests, diagnostic tests, progress tests, and achievement tests, and the benefits of CRTs for the students, teachers, and curriculum.	Teachers, curriculum developers, researchers, and language teachers can benefit from this study, as it can help them understand the top issues in and benefits of CRTs.	4, 9	1	21	15	4
Brown (2004a)	This study reviews issues in task-based testing and performance testing and presents a history of performance assessment and overviews the global and specific trends in related literature.	The study can provide testing and performance assessment knowledge to teachers, testers, and scholars in the field.	1, 2, 4, 5, 7, 9, 10, 11, 12, 16, 18, 25	1	21	15	1, 2, 3, 4
Brown (2004b)	This study describes and clarifies the Yates' Correction Factor which is used with chi-squared analysis under certain conditions by reviewing a definition of chi-squared analysis, as well as the use and application of Yates' Correction.	Researchers in the field need to be familiar with issues related to the Yates' Correction, and to do this, they need to know about the regular chi-squared test, too.	25	1	21	15	16
Brown (2004c)	This study reviews three issues: the reason why students intentionally perform poorly on tests, factors needed for the interpretation	This study is informative to teachers helping them deal with the motivational factors related to	1, 2, 4, 9, 10	1	21	15	1, 4, 16

	of gain scores, and the strategies for countering such factors.	test taking and thinking about reasons for students' performances.					
Brown (2004f)	This study examines the general issues of test fairness, test bias, standard British/American differences, Englishes in testing, i.e., multiple Englishes in a single EFL/ ESL test, and English language proficiency. Much wider discussion of proficiency tests is covered in Brown (2019d, 2021b).	Here, scholars are invited to conduct further study into the complicated relationship among purpose, test bias, and the various Englishes of the world.	1, 2, 4, 7, 9, 15	1	21	15	1, 4
Brown (2004g)	This study introduces, reviews and evaluates nine quantitative and statistical research books in applied linguistics and compares them in terms of their overall features, and statistical and conceptual themes.	Graduates, scholars, and postgraduates can consider using the books introduced and reviewed here for doing research, as they are highly informative.	2, 22, 23, 25	1	21	15	1, 4, 16
Brown (2005a)	This paper examines three issues in relation to the nature and usefulness of G-studies, D-studies, and the differences between them, as well as the time needed to use them in analyzing data.	Researchers can use G theory to solve various types of measurement problems in language testing, and research.	2, 7, 25	1	21	15	1,3,4, 16
Brown (2005c)	This article reviews the definition of research and the characteristics of quantitative research, especially, reliability, validity, replicability, and generalizability.	Researchers can use the content of this article to enhance their understanding of quantitative/qualitative research	1, 2, 11, 23	1, 2	21	15	1, 3, 4, 16
Brown (2005d)	This article explores the reasons for publishing, tips for publishing articles and books, and the steps required in publishing journal articles and books.	This study can help teachers. and researchers understand and participate in publishing journal articles or books.	9	1	21	15	4
Brown (2005e)	This study reviews various aspects of language testing including test purpose (CRTs/NRTs), options (selected-response, constructed-response, personal-response tests), washback issues and language testing constraints (political, functional, and economical).	The study is a clear review of the four issues discussed, so teachers, teacher trainees, language testers, and researchers can all benefit from it.	4, 9, 11	1	21	15	4, 5

Brown (2006e)	This study explains a number of research issues including sampling and generalizability with a main focus on samples and populations, random samples, stratified samples, transferability and adequate generalizability.	Researchers should be careful about making statements about large populations on the basis of small samples or samples of convenience.	1, 11, 25	1	21	15	1, 3, 4
Brown (2006h)	This study provides helpful tips for those interested in taking a language testing course and guidance for checking out Internet websites, subscribing to one or more language testing journals, joining a language testing organization, and reading recent books.	The study is useful for those who are interested in taking a language assessment course and provides them with many websites needed for starting and continuing the study of language testing.	5, 9	1	21	15	1-17
Brown (2007a)	This study examines sample size and power and discusses issues related to null hypotheses, Type I and Type II errors, definition of power, errors resulting from ignoring power, and calculation of power.	L2 researchers need to consider both Type I and Type II errors; Also, they need to notice power statistics, because they help understand Type II threats to our studies.	2, 11, 25	1, 2	21	15	1, 16
Brown (2007b)	This study discusses sample size and statistical precision including samples and populations, statistics and parameters, statistical precision, descriptive and inferential uses of statistical precision, and the relationship between sample size and precision.	Researchers can benefit from studying the present article as it will help them understand the notion of precision and its relationship to large sample sizes.	11, 25	1	21	15	16
Brown (2007e)	This article examines a number of aspects of writing test score consistency using both classical theory and generalizability approaches that can help in improving the consistency of institutional scoring and testing procedures.	Researchers should not use reliability at the expense of validity; they need to consider both, as both can lead to desirable consequences and responsible test use.	1, 2, 4, 25	2	12	1, 2, 3, 9, 10	1, 7
Brown (2008a)	This study explores the Bonferroni adjustment with reference to the problems with interpreting multiple statistical comparisons, the probability of one or more *t*-tests being spuriously significant, and	Researchers need to know that in addition to the Bonferroni adjustment strategy, they can use the ANOVA family of statistical tools.	25	1	21	15	16

	ways to solve these problems by using the Bonferroni adjustment.						
Brown (2008b)	This study clarifies the nature and definition of partial eta squared and what partial eta^2 measures, what other forms of eta^2 readers should know about, and how a partial eta^2 value of .29 should be interpreted.	There are implications for researchers, as they can use the article to clearly deeply understand a number of germane issues related to ANOVA studies.	25	1	21	15	16
Brown (2008d)	This study examines issues related to the procedures, steps, purposes, uses, constraints in doing testing-context analysis, and explains related issues such as stakeholder friendly curriculum, needs analysis, construct validity and defensible testing, and the components of curriculum.	Language testers, teachers and researchers can benefit greatly from this article because language testing is and should be an important component of language curriculum development.	1, 2, 4, 5, 7, 9, 11, 17, 19, 22, 25	1	21	15	1, 3, 4, 5, 6, 16
Brown & Bailey (2008)	This article compares the characteristics of basic language testing courses studied in the years 1996 and 2007 in terms of the instructors, course characteristics, and students' views in terms of both differences and similarities.	The study lists language testing topics which can serve as a source of ideas for those designing or revising language testing courses.	1, 2, 3, 4, 9, 19, 25, 28	1, 2	14	1, 2, 3, 4, 5, 8, 9, 17	1, 4, 16
Brown (2009f)	This study reviews the mistakes Brown made and the problems he faced in language curriculum development including his beliefs and assumptions over 34 years.	The study has implications for researchers, teachers, and curriculum developers. Brown admits his mistakes which is a big lesson in itself.	5, 9	1	21	15	4
Brown (2009b, 2009d, 2009e, 2010a, 2010b)	The difference between principal component analysis and exploratory factor analysis and related rotations is reviewed (2009b, 2009d), as well as item/subscale analysis and the relative proportions of total, reliable, common, unique, specific, and error variances (2010a) are explained. Finally, using factor analysis to reduce the number of variables in a study and supporting the relationships among variables (2010b) are described.	For running factor analysis, researchers can benefit from the differences mentioned between PCA and EFA. They need to consider the other purposes factor analysis can serve. Also, they are shown what happens when both oblique and orthogonal rotation methods are used.	1, 25	1	21	15	1, 16

Brown (2011a)	This study examines Likert-items and scales of measurement and guides the readers in how to analyze and treat "Likert-scale" items on questionnaires as nominal, ordinal, interval, or ratio scales and how to design such items.	Language researchers and teachers can use the information in this article to effectively design Likert-like items.	2, 25	1	21	15	1, 4, 16
Brown (2011b)	This study examines 44 generalizability theory studies in terms of the relative magnitudes of the variance components; the results explore patterns in the relative contributions to test variance of various individual facets and interactions among them for different types of tests.	Language testers can use the results to inform their test development strategies.	2, 3, 4, 7, 9, 11, 12, 19, 23, 25,	1, 2	21	9, 10	7, 8, 9, 10, 11, 12, 13, 14
Brown (2011d)	This study examines the uses and differences between confidence levels, limits, and intervals in language testing to help in calculating and understanding standard error statistics.	Testers and researchers can use the confidence intervals, limits, and levels for interpreting standard errors.	25	1	21	15	16
Brown & Ahn (2011)	This study examines four types of instruments for testing L2 pragmatics with use of Generalizability theory and multifaceted Rasch (FACETS) analyses and tackles the relative severity of individual raters, and difficulty of item types, characteristics, and functions, as well as the effect of the five-point scale on each test.	Testers and researchers can use this discussion of issues related to item types, functions, and characteristics as well as the numbers of different raters for the purposes of maximizing test dependability.	2, 3, 4, 5, 9, 12, 18, 25,	1, 2	13	2, 10, 16	1, 16
Housman et al. (2011)	With use of a standards-based assessment tool and an oral language proficiency rubric, the project develops a comprehensive oral language proficiency assessment to collect data needed for examining Hawaiian language immersion program students.	The study is of potential use for testers who are interested in developing a program-based oral proficiency assessment for learners at various levels of language proficiency.	2, 3, 9, 10, 25	2	13	1, 6, 8, 9	1, 4, 8
Brown (2012h)	This study discusses key issues in statistics such as distributions, assumptions, statistical significance, meaningfulness, multiple statistical tests, causal interpretations, null results, and the ways they should be treated in L2 research statistics.	One of the key implications is that significance does not indicate meaningfulness; statistical significance and meaningfulness are different things.	2, 11, 25	1	21	15	16

Brown (2012i)	This study reviews the agreement coefficient and the Kappa coefficient and describes how to calculate rater/coder agreement and Cohen's Kappa by presenting simple and clear examples.	The study can help researchers understand how to lay out the data and calculate agreement and Kappa coefficients.	2, 25	1	25	15	16
Brown, Janssen et al. (2012, 2019)	The relationships between readability indexes and cloze passages and estimating readability using cloze passages were examined based on data from Russian university English language students.	The analyses can significantly contribute to researchers' understanding of the relationship between readability and cloze passage performance.	1, 2, 3, 4, 25	2	24,	1, 6	1, 10, 16
Brown (2013a)	This article examines the whole body of cloze testing research conducted by JD Brown over the course of 25 years in terms of the questions raised, answers, and results.	The study has implications for researchers on how research progresses and for teachers who want to use cloze tests for assessing their learner' performance.	1, 2, 3, 4, 25	1, 2	21	15	1, 3, 4, 16
Brown (2013b)	This study provides solutions to problems with classroom testing and criterion-referenced testing and deals with issues such as test writing practices, test development practices, and test validation practices.	Teachers can use the findings of the study for the development of tests and test items for achievement purposes in their classes.	1, 2, 3, 4, 5, 6, 25	1	21	15	1, 3, 4
Brown (2013h)	This study examines likelihood ratios, continuity-adjusted and Mantel-Haenszel chi-squares with a view to calculating simple chi-square for a 2 × 2 contingency table, checking the assumptions of Pearson's chi-square, and using variations on the chi-square family of statistics.	The study can help researchers and language testers understand the key concepts germane to chi-square type statistics for data analysis in their research.	25	1	21	15	16
Brown (2013i)	This study examines teaching statistics in language testing courses, including potential approaches and classroom tools to help overcome statistics anxiety. It also includes ideas for topics and advice for teacher trainers.	Teacher trainers can use the article for ideas and strategies to use in teaching statistics in language testing courses.	1, 2, 3, 4, 9 25,	1	25	15	4,16
Brown (2014b)	This study reviews the differences between NRTs and CRT-referenced families of tests, the strategies employed for the development and validation of NRTs and CRTs, and the	Since the study provides a concise account of the nature CRTs and NRTs, it can be of potential use for researchers, testers, and teachers	1, 2, 3, 4, 25	1	25	15	1, 4

	differences in NRT and CRT development and validation strategies.	to help them understand, use, or explain these concepts.					
Brown (2014c)	This article examines world Englishes and language testing, as well as the ways language testers treat world Englishes; it also explores the concepts of inner-, outer-, and expanding-circles of English(es).	This study has implications for testers and researchers and should inspire them to do much more research on WEs, ELF, and EIL, in relation to language testing.	1, 3, 9, 19	1	25	15	4
Brown & Alaimaleata (2015)	This study investigates the validity and reliability of the Samoan Oral Proficiency Interview. His co-authored article on the reliability and validity of the Arabic Proficiency Test (Brown & Hachima, 2005) is also informative.	Teachers of Samoan and Arabic can use these tests for assessing their learners' spoken proficiency because they have been shown to function well.	1, 2, 4, 5, 9, 25	2	16	1, 2	1, 2, 3, 8
Brown (2016a)	This article is a professional reflection on Brown's forty years of research and investigation in applied linguistics and focuses on his formative professional development.	Readers can learn much from Browns' experiences in language testing, research, quantitative research methods, curriculum and program evaluation, and development of research topics.	4, 9, 25	1	21	15	17
Brown (2016d)	This study defines the notion of research and the characteristics of mixed-method research, and discusses qualitative and quantitative research issues; it also elaborates on various forms of legitimation.	The study is of potential use for those interested in doing mixed-method research, and so it could serve as a sound reading in a language research course.	1, 11, 23	1	21	15	1, 4, 5
Brown (2016f, 2017c)	These studies examine the notions of internal and external reliability in quantitative research and reliability of NR/CR tests and consistency in research design in terms of categories and subcategories.	Understanding the difference between external reliability and internal reliability can help testers and researchers to perform, evaluate, and interpret L2 research.	1, 2, 4, 25	1	21	15	1, 4
Brown (2010c, 2017a)	These studies detail JD Brown's professional development, fundamental mistakes, and crucial lessons, as well as discussing language testing and research method issues, paradigms, processes, and challenges in applied linguistics. Some of his most important personal changes appear in Brown (2000f).	These articles provide a window into the ways JD Brown developed professionally. It also provides some interesting and significant information about doing language testing and research.	3, 4, 7, 9, 10, 17, 25, 26	1	21	15	4, 17

Brown (2014e, 2017b)	These studies describe various types of rubrics used for assessing either oral or written language and the primary differences between analytic and holistic rubrics, as well as the importance of rubrics in general.	Teachers can use the study in assessing and scoring the written and oral language output to help them rationally select and create holistic or analytic rubrics.	2, 4, 10, 12, 17	1	21	15	1, 4
Purpura et al. (2015)	This study examines quantitative research in applied linguistics and examines the use of measurement instruments and scores, validity and validation, and presents a comprehensive framework for score interpretation and score use.	The study is of potential use for researchers, as they can use the brief checklist for quantitative data collection and for suitable data treatment in future analyses.	1, 2, 3, 4, 12, 23	1, 2	25	15	1, 4
Brown (2018a)	This article examines the reliability of dictation tests whether or not K-R21 can be used effectively. The results indicated its effectiveness.	Test designers can learn how to use K-R21 for estimating the reliability of dictation tests cautiously and together with other methods.	2, 4, 7, 25	1, 2	21	1,2, 3, 15, 16	1
Brown (2019a, 2019b)	These studies review the effectiveness of feedback and assessment in classrooms (Brown, 2019b), modes of feedback (teacher-feedback, self-feedback, peer-feedback, and conference-feedback, or presentation-feedback), tools for giving feedback and strategies making feedback effective (Brown, 2019a).	The studies are helpful for teachers and teacher trainers. They can use the content of the articles to concisely learn the kinds of feedback available and the steps that can be taken to apply feedback.	5, 9, 19	1	25	15	4
Brown (2023)	This article reflects on twelve sources in the literature that influenced Brown's thinking on connected speech and how they influenced him. It also provides a number of bonus take-aways about things he learned professionally in the process.	Scholars can learn much about developments in connected speech from these 12 connected speech resources, as well as from the authors' professional reflections.	1, 3, 2, 4, 9	1	21	15	1, 4, 9, 13

Table 3

Analysis of Book Chapters

Book Chapters	Briefing	Implications
Brown (1983b)	Here, Brown reports two studies concerning cloze test about: the validity of the cloze test and the reliability of the cloze test. Considering various types of reliability, he believes that the sole focus on the validity of cloze test is not enough, as reliability is also important for cloze tests.	These studies have implications for test designers and researchers who want to understand the theoretical issues and empirical findings regarding cloze test validity and reliability. Teachers also need to read this chapter in order to be aware of ways to develop cloze tests.
Brown (1984a)	This study examines the effects of differences in samples on cloze reliability and validity, test characteristics, and scoring methods, and the fit of cloze tests to samples, as well as the strengths of relationship between ranges of ability and cloze test reliability and validity.	The study has implications for language testing specialists, as it covers specific issues in psychometric theories. It also has practical implications for language teachers, who may need to pretest any cloze test before administering it for norm-referenced purposes.
Brown (1989a)	This chapter explores program evaluation in educational psychology and deals with issues such as the differences and similarities between testing, measurement, and evaluation, while synthesizing historical trends and program evaluation in terms of formative, summative, product, process, qualitative, and qualitative dimensions.	Teachers can use the chapter to improve their evaluation-related process, procedures, data-gathering, and assessment. This can help them to evaluate their curriculum and their own teaching processes and results, all of which can lead to meaningful learning.
Brown (1989f)	This study examines the listening needs of the students at the University of Hawai‘i through the development of systematically designed listening curriculum. The results argue for providing targeted listening materials for meeting learners' communicative needs.	In teaching listening, teachers should first conduct needs analysis and then, based on the results, provide students with either graded or authentic listening practice, all the while adapting the materials to the level and needs of learners.
Brown & Pennington (1991)	This study reviews the definition of evaluation, redefines it, and provides a brief account of procedures for language program evaluation using various categories of information including existing records, tests, observations, interviews, meetings and questionnaires; it also elaborates on the role of program administrator and finally reviews the implementation of program evaluation processes.	The study is of potential use for course administrators and program managers, who can use the findings for collecting data, making decisions, and evaluating their programs. Also, the chapter can benefit and inspire other language professionals to do program evaluation.
Pennington & Brown **(**1991); Brown (2000e); Palacio et al. (2016)	These studies discuss the definition of curriculum, the role of administrators in curriculum, and a curriculum process model, including needs analysis, objectives, testing, materials, teaching, and evaluation of curriculum, as well as testing purposes (Brown, 2000e), and aligning language testing and curriculum (Palacio et al.,2016).	The studies provide effective definitions and models for all stakeholders including teachers, program managers, and administrators. With use of the theoretical foundation laid down in these chapters, language professionals can create a curriculum and appropriately use both CRTs and NRTs.

Brown (1993a)	This study discusses theoretical issues in language curriculum, describes examples of social meaning in curriculum, and advocates making changes through curriculum by describing a curriculum process model, a research and development model, a social interaction model, wherein curriculum development is a political process.	Teachers can appropriately analyze learners' needs in terms of tasks to be included in the syllabus and the kinds of syllabuses they need. Lesson to draw: in developing curriculum, the practical, political, and innovative issues all need to be taken into account.
Bailey & Brown (1996)	This chapter designs, develops and examines a Likert-scale questionnaire for the purpose of tapping the structure, content, and attitudes of students towards introductory language testing courses and the relationship between language testing and language teaching.	The chapter can be of use to teacher trainers who teach language testing in pre/in-service courses. Also, the questionnaire is appended so it can be used or adapted by researchers for future studies.
Brown (2001a)	These studies investigate six types of pragmatics tests in two settings: an English as a foreign language setting and a Japanese as a second language setting. The Japanese translations of the six tests worked much better than the original English-language versions. However, the latter were argued to be of much use, too. Further examination of these issues is provided in Brown (2000d, 2001g, 2018c).	Since testing characteristics can change depending on the local context, teachers can use EFL tasks associated with six pragmatics tests targeted to the local needs, interests, preferences, and purposes of the learners. The implications are of statistical use to test developers, too.
Brown, Cunha et al. (2001)	A Portuguese version of the Motivated Strategies for Learning Questionnaire was developed and studied in terms of its construct validity and reliability. Also, the new version was administered in private and public universities and differences were found in the performances of the learners in the two settings.	Instructors can use the questionnaire for detecting the kind of cognitive strategies Portuguese learners use. Also, it can be translated for research and be tailored to learners' needs and interests in other contexts.
Brown, Robson et al. (2001)	This study of Japanese EFL students covers various issues related to individual differences in terms of personality, motivation, anxiety, and language learning strategies which affect learning as well as the language proficiency of the learners.	Teachers can use the findings to consider various types of motivational profiles, anxiety-types, aspects of learning strategies, and proficiency levels in adjusting their language teaching for learners in Japan.
Brown (2003e)	This chapter examines central issues in language testing and presents a comprehensive assessment system for language programs. As such, it also discusses practical and theoretical issues prerequisite to familiarity with language testing.	The study is of potential use for language testers, teachers, and administrators, as it includes practical information about integrating testing into course syllabuses and developing sound curriculum.
Brown (2004d)	This chapter examines the scope of, characteristics of, and options in applied linguistics research, beginning with a definition of applied linguistics research, and then elaborating on traditions, roles, problems, validity, generalizability, and transferability as the qualitative and quantitative research standards for sound research.	The chapter is of potential value for both experienced and novice scholars and professionals; it provides a solid background for the ins and out of research in applied linguistics. Also, it could usefully be included in research-related course syllabuses.
Brown (2004e)	This chapter examines standardized tests and their characteristics and uses, as well as tendencies toward grade inflation.	The study can help language testers and researchers in the field of language testing to develop responsible standardized tests and help teachers to accurately think about responsible grading.

Brown (2006d)	This chapter describes six major categories of curriculum activities (underpinnings, contexts and organization, gathering information, outcomes, and wh-questions) including a total of 15 curriculum facets and almost 100 individual subparts.	Teachers, scholars, professionals, and course managers can use the chapter to understand the theoretical and pedagogical knowledge base related to curriculum development.
Brown (2008c)	This study presents the statistical analyses required for improving pragmatics tests with use of classical theory and generalizability theory approaches. Conducted in a Korean context, first a generalizability study (G study) was conducted, followed by a decision study (D study).	The chapter can be of value to graduate students, postgraduates, and researchers. The researchers in ELT field can especially benefit from observing how the author applied G theory in two steps: a G-study and then a D-study.
Brown (2009a)	This chapter covers crucial issues related to needs analysis including the nature and purpose of needs analysis, the literature on needs analysis, steps and procedures for doing needs analysis (considering both quantitative and qualitative research), data collection in needs analysis, interpreting results, and reporting needs analysis research.	Since the chapter covers theoretical issues in needs analysis research with tangible examples and with clear stages and steps for performing needs analyses, it may prove very useful for teachers, researchers, program managers, administrators, and curriculum developers in doing needs analysis.
Brown (2009c)	This chapter first presents some pre-reading questions, then provides a comprehensive account of open and closed response items on questionnaires, and includes discussions of questionnaire use, item types, and the kinds of information they can obtain.	The chapter is of potential use in classrooms and can be included as part of syllabus. Due to the clear examples supplied in the chapter, it can help readers understand the nature of such items and questionnaires.
Brown (2009g)	This study surveys issues related to using spreadsheet programs by defining what spreadsheets are and examining their use for recoding, organizing, and understanding classroom-based assessment data.	Spreadsheet programs can be of potential use to classroom teachers in assessing, keeping records, and grading students.
Brown (2011c)	This chapter reviews the history of quantitative research in L2 studies, its current status, and its future directions. It also elaborates on quantitative research in terms of its nature and provides guidelines for doing such research. It also reviews the books available on this topic.	The chapter can be used by graduate students, researchers, and research instructors, as it provides background and information about how to do quantitative research. It also presents ideas for future research topics.
Brown (2012c)	This chapter examines various issues in language testing including classical test theory validity, classical test theory reliability, consequential validity in terms of criterion/norm referenced testing, score consistency, test content, test items, and learners' performance.	The chapter will of use for language teachers, language tasters, program managers, and researchers in helping them analyze test items and the effectiveness of classroom tests.
Brown (2012d)	This chapter examines concerns relevant to choosing the right type, function, and purpose of assessment so that tests can be tailored to curriculum objectives and assessment-based decision making. Test types and purposes compatible with decision making are detailed.	Teachers need to relate the findings of language testing to language pedagogy in the classroom and realize that for each purpose a particular kind of test should be used.
Brown (2012e)	This chapter examines the history of classical test theory (CTT) and related issues such as the relationship between observed scores and true scores, and the factors affecting these scores. Thus, it details the main elements of CTT.	The chapter can initiate specialists and non-specialists into the ins and outs of CTT. It clearly explains the elements of CTT, such as observed score variance, true score variance, and error variance, which are all central to CTT.

Brown (2012f)	This chapter provides an insightful account of questionnaires as written instruments, including types, items (e.g., open-ended responses), and uses of questionaries as well as the process, procedures and data collection issues relevant to questionnaires.	The chapter can be of use to novice and experienced researchers by making them aware of the process and procedures involved in questionnaire-based research.
Brown (2012g)	This chapter explores and compares the principles of traditional and English as an international language (EIL) curriculum development and covers issues such as the target language and culture, reasons for learning English, curriculum delimitations, and the basic units for analysis, selection, and sequencing the curriculum.	The chapter is of potential benefit for teachers, researchers, materials developers, and curriculum developers, who can use the procedures suggested for developing EIL curriculum and doing EIL research.
Brown (2012j)	This study examines the processes involved in writing up replication reports including issues such as the way to do replication studies, the content of such a research paper, the kinds of original study contents to include, and ways to report findings.	The chapter is of potential use for those who would like to do replication research suggesting ways to do it, what to include from original investigation, and how to deal with exact, approximate, and conceptual replications.
Brown (2013c)	This study provides a brief discussion of criterion-referenced tests and norm-referenced tests, and then, it covers statistical issues such as entering the data, pre-test/post-test practice effect, and running, reporting, and interpreting a one-way repeated-measures ANOVA.	The chapter is useful for both experienced and novice researchers, as it covers issues germane knowing how to treat statistical analysis in running, reporting, and interpreting ANOVA.
Brown (2013d, 2016e)	Brown (2013d) reviews four sets of issues: item banking, technology and computer use, computer-adaptive language testing, and the literature, content and delivery issues on computer-based language testing. Brown (2016e) also discusses the use of technology in language testing, including information from key articles from over two decades.	The findings of these studies have implications for researchers, testers, and teachers, as it provides background on computer-based instruction and assessment in terms of different ways to assess language skills and subskills.
Brown (2013e)	This chapter covers cut scores and standard setting in terms of the steps and procedures related to standards setting, the options and methods for standards setting, errors involved in standard setting, as well as cut scores and reliability and dependability.	The study has implications for researchers and language testers working in the field of research statistics. They can use the findings to help in setting standards and using cut scores for both criterion-referenced and norm-referenced tests.
Brown (2013f)	This chapter describes a three-part chain of inference (a conclusion resulting from evidence or logic) and inferring (the process leading to the evidence or logic) including elaboration of constructs from variables, populations from samples, and probabilities using inferential statistics.	Researchers and testers doing quantitative studies can benefit from knowing about the three-part chain of inferences explained in this chapter so they can choose the right strategies for their data and the purpose of their study.
Brown (2013g)	This chapter investigates sampling, the difference between populations and samples, probability versus nonprobability sampling, methods of probability sampling, methods of nonprobability sampling, and choosing the right sampling strategy.	Researchers need to know the different sampling options they have to create a sample representative of their target population. Also, they need to consider participant attrition and think of the possibility of incomplete data.
Brown (2013j)	This chapter examines test score dependability and decision consistency drawing on G-theory, estimating NRT score error, calculating signal-to-noise ratios for NRTs, threshold loss agreement, squared error loss agreement, phi	The chapter can be incorporated into the syllabus for any advanced language testing course as it will initiate teaching professionals and researchers into important advanced testing statistics.

	dependability, the standard error for absolute decisions, and signal-to-noise ratios for CRTs.	
Youn & Brown (2013)	This chapter reviews issues in the testing of L2 pragmatics and language testing and provides effective background for understanding pragmatics-related tests.	The chapter can be of potential use for testers, researchers, graduates, postgraduates, professionals and teachers, as it provides a brief review of key issues in L2 pragmatics.
Brown (2014a)	This chapter investigates the literature and basic logic for classical theory (CT) reliability, the relationship between norm-referenced tests and reliability, and various approaches to reliability, as well as error-estimation approaches to CT reliability.	The chapter has implications for language researchers and testers who need to understand and use the practical and theoretical aspects of CT, especially for deciding which CT reliability strategy is suitable for their test and its purpose.
Brown (2015a, 2015b, 2015c)	These chapters examine the advantages and disadvantages of advanced quantitative research (Brown, (2015a) and the characteristics of sound research and research methodology (Brown, 2015b, 2015c)	These chapters can help researchers, testers, and others interested in statistics understand the characteristics of sound research and advanced research methods.
Brown (2016b)	This chapter describes 12 assessment formats grouped into four categories: productive-response (short-answer, fill-in items, and performance assessment); individualized- response (i.e., dynamic assessment, continuous, and differentiated); receptive-response (multiple-choice, true-false, and matching items); and personal-response (conferences, portfolios, and self/peer assessment).	The chapter describes a variety of testing format options for teachers and testers to choose from in matching their assessment tools to the applicable language pedagogy in order to provide positive washback on teaching and learning processes and outcomes.
Brown & Trace (2016)	This chapter examines assessment issues related to planning, designing, and performing assessment and introduces teachers to various item types in four categories: selected-response items, productive-response items, personalized-response items, and individualized-response items. For further explanation about determining cloze item difficulty, see Trace, Brown, et al. (2017).	Teachers, testers, and researchers can benefit from understanding the importance of planning, and carefully designing classroom tests and realizing the significance of feedback after the assessment has taken place.
Brown, Trace, et al. (2016)	This chapter investigated item-level data from fifty 30-item cloze tests randomly administered to university-level examinees. Fairly large sample sizes were gathered in two countries: Japan (N=2,298) and Russia (N=5,170). The results revealed that different items were functioning well for the two nationalities.	The study can serve language testers and language teachers who need to develop cloze tests and can add to their own theoretical knowledge base associated with cloze test concerns and issues.
Brown (1984b, 1993e, 1995j, 2001i, 2018b)	This chapters discusses language testing explaining norm-referenced tests (aptitude, proficiency, and placement testing) and criterion-referenced tests (diagnostic, progress, and achievement testing) and various test development issues.	These chapters can help ESL/EFL teachers understand language testing and help them develop norm-referenced tests and criterion referenced tests for decision making that will serve their learners' needs.
Brown (2018d)	This chapter examines cloze testing for proficiency or placement purposes and reviews fixed deletion cloze patterns (including open-ended every seventh word scoring for exact answers or acceptable answers, every second word, multiple-choice cloze, cloze elide, and C-test) and rational deletion cloze.	Teachers and novices to the field of testing and language education can benefit from the discussions in this chapter and learn how to perform the five steps for selecting a text or passage and developing a cloze test for classroom purposes.

Brown (2019c, 2020)	This chapters discusses the literature on Global Englishes, English as a lingua franca, and English as an International Language from the perspectives of problematic issues in language testing (linguistic norms, testing cultures, test design, testing processes and testing in various contexts), the native-speaker standards and models, and the international standardized English language proficiency tests.	The chapters can be of potential use for teachers, researchers, and graduate students by, among other things, helping them come up with insightful research questions associated with World Englishes and language testing and providing them with the theoretical knowledge base germane to proficiency tests and Global Englishes.
Brown (2022a)	This study zeros in on the significance of and reasons for conducting needs analysis in second language classrooms. It elaborates on tools, sources, and procedures for performing needs analysis.	Teachers can use the practical tips and questionnaire examples to help them understand learners' needs, interests, and preferences, and diagnose learners' weaknesses and strengths.
Brown (2022b)	This chapter addresses Chinese language native-speaker-ism, the impossible dream of expecting learners of Chinese to become near-native speakers, and how to set goals using Chinese for specific purposes and the related curriculum and needs analysis.	The chapter has implications for teachers and learners of Chinese; it can help them sort out their reasons for teaching and learning Chinese, and do needs analysis and the curriculum development.

Table 4

Analysis of the Books

The Books	Briefing	Implications
Brown (1988c)	This book deals with statistical terms and concepts, the organization of statistical research reports, the system of statistical logic, and how to decipher tables, charts, and graphs as well as the skills necessary for understanding statistical research in language learning.	The book was originally written for readers with no previous statistical competence or experience. Therefore, it can be used as a coursebook providing the readers with the skills for making judgments about the value of the results of a study.
Hudson et al. (1992)	This book deals with various issues such as the role of pragmatics and its contrastive realization in communicative competence, social distance, relative power, the causes of pragmatic failure, and variables in speech act realization in Japanese and American contexts.	Serving as a generic approach, the framework can be applied to contexts beyond America and Japan and help teachers to develop their theoretical and pedagogical knowledge base for assessing learners' performance on the basis of speech acts.
Brown (1995a, 1995g)	*This comprehensive book* covers the building-block elements for language curriculum development. The book details theory and practice in relation to needs analysis, goals and objectives, testing, materials, teaching, and program evaluation. Overall, he advocates relating language testing closely to curriculum pedagogy.	Teacher trainers can use this book as a coursebook for designing and developing a curriculum and meeting learners' needs and interests. Also, experienced and novice scholars can use it as a reference point for curriculum research.
Brown & Yamashita (1995b)	This book is a collection of articles *covering* issues such as norm/ criterion-referenced tests, cooperative assessment, assessing young learners, uses of TOEIC and TOEFL, washback, oral proficiency, non-verbal ability, cloze testing, and pronunciation validity.	Instructors and scholars can use this collection as course papers and as potential sources for research purposes. Also, it can be used as an effective source for following and enriching professional development in language testing.

Hudson et al. (1995)	The book presents phase two of the instrument development process and multiple methods for assessing cross-cultural pragmatic abilities. It therefore covers various issues such as classification of test methods, variable distribution across tests, development of the discourse completion test, item specifications, piloted instruments, analysis of piloted instruments, pragmatics issues, and speech act strategies.	Since the study uses both quantitative and qualitative approaches for the development of prototypic instruments, it can be useful and informative for test developers, test users, and scholars working in the field of test instrument development, and it provides valuable insights into the steps and processes involved in test and instrument development.
Brown (1996a)	This book deals with the development and adaptation of different kinds of language tests. In general, it covers issues such as test types, test development, use and improvement of tests, description of results and score interpretation, test reliability and correlational issues, and test validity, standards, and testing in language curriculum.	The book will be instructive to test designers and developers, test users, scholars, teachers, and administrators. Due to the comprehensiveness of the book in terms of language testing, the book can be used for program level decisions or curriculum level decisions.
Brown (1998a, 2013k)	These books are the first edition and much revised second edition. They describe numerous assessment activities for EFL/ESL classes and procedures for performing real performance assessment are examined. Also, they cover key issues such as alternative assessment, conferences, logs, journals, and portfolios, assessment scales, self-and peer assessment, alternative and feedback perspectives, and alternative ways for grouping learners for assessment.	The book will be of use for the graduate students, undergraduates, postgraduates, teachers, and scholars as part of a course or as a self-study book full of ideas for language classroom assessment.
Norris et al. (1998)	Aimed at providing guidelines for performance assessment, the book covers various issues such as alternative assessment, alternatives in assessment, performance assessment, needs analysis, task-based performance reliability, validity and assessment, test/item specifications, and item prompts.	Teacher trainers can use the book as a coursebook for language testing or include chapters from it in their syllabuses. Since the book covers a wide range of issues on modern and new approaches and procedures for language testing, it has potential those holding workshops, too.
Brown, Hudson, et al. (1999)	Korean performance assessment and testing Korean as a foreign or second language are investigated with reference to task-based performance assessment, item/test specifications, selection, revision and validation, and dissemination on the internet.	The book will prove useful for teachers, testers, and program managers in Korean language contexts, as it provides strategies for viewing and implementing performance assessment.
Iwai et al. (1999)	This booklet is a report on the results of an on-going curriculum development needs analysis aimed at creating performance-based tests for Japanese language courses at the University of Hawai‘i at Mānoa.	Teachers, learners, curriculum developers, and program managers should consider a needs-analysis-based approach as an integral part of every educational agenda which in turn can affect the cycle of learning, teaching, and assessment.
Hudson & Brown (2001)	Containing eight research studies, this book examines alternative approaches to test development and covers issues such as evaluating nonverbal behavior, collocational knowledge and L2 vocabulary, pragmatic picture response tests, a three-phase pragmatic performance assessment, revising cloze tests, task-based EAP assessment, and criteria-referenced tests.	The results of these studies may prove useful for scholars and test designers in that they can learn from the various research designs and test development projects. Since the book also contains elaboration on some non-standard types of language assessment, it may also prove useful for nonexperts and language teachers who use tests for classroom purposes.

Brown (2001e)	This book contains six chapters and covers a wide range of topics including planning and designing a survey project and instrument; gathering, compiling, and analyzing survey data; analyzing survey data qualitatively; and reporting on a survey project. Each of the topics includes extensive samples.	Teachers, administrators, and researchers may find this book useful due to the plentiful examples, careful definition of terms, applications exercises, review questions, appendices, and suggestions for further reading which all make the book more engaging and hold the interest of readers.
Brown (2002a); Brown, Hudson, et al. (2000)	Authored by expert researchers, the book deals with issues on test development, methodological and statistical issues, task-dependent scaling, and assessment needs of language learners. It also operationalizes the instruments and procedures associated with task-based performance assessments. Also see, the Brown, Hudson et al. (2000) investigation into performance assessment of ESL and EFL students which is complementary to this work.	Teachers can learn from this work how to conduct performance assessment on both receptive and productive skills rather than sacrificing one to the advantage of the other. Also, the appendices (one third of the book) at the end of the book can serve as sources of assessment ideas for researchers and teachers alike.
Brown & Hudson (2002)	Comprising seven chapters, this book examines alternate paradigms, curriculum-related testing, CRT items and item statistics, reliability, dependability, the unidimensionality of CRTs, test administration, feedback giving, score reporting, and the validity of CRTs.	The book can help teacher trainers and language teachers who lack background technical knowledge on CRTs and language testing because it covers strategies for pedagogical decision-making in testing-driven instruction.
Brown (2005b)	Comprising 11 chapters, this book examines various topics such as types and uses of language tests; adopting, adapting and developing language tests and test items; item analysis; describing results; interpreting scores; correlation; test reliability and dependability; validity; and also using tests in real situations.	The book is recommended as a coursebook for graduate students, undergraduates, postgraduates, and language teachers. Also, researchers, test designers, and administrators may find the book useful as it covers both theoretical issues and practical tips and techniques for using tests in classrooms.
Brown & Kondo-Brown (2006); Brown (2012a, 2023)	These books are about connected speech (CS) and new ways for teaching and assessing CS in EFL/ESL contexts. Also, a book chapter by Brown and Trace (2018) further examines CS dictations for testing listening and reduced-forms dictations are also studied in Brown and Hilferty (1998); more recently, Brown (2023) reviews and reexamines the connected speech issues at length.	Teachers need to cover pronunciation in more depth than they do traditionally by teaching sounds as the occur in connected utterances in the form of connected speech rather than teaching in phonemes in isolation.
Brown & Rodgers (2003)	The book covers the doing of second language quantitative/qualitative research by describing the processes for research design, data analysis, and report writing and providing plenty of activities and example mini-studies	The book can be used as a coursebook, or as a part of syllabus for a research methodology course. So, it could be of potential use for the graduate student, post-graduates, and scholars.
Brown, Davis, et al., (2012)	This study includes the validation, linking, and use of scores on the upper-level EIKEN examinations for the purpose of predicting the Test of English as a Foreign Language (TOEFL) Internet-based Test (iBT) scores. The two test batteries appeared to be measuring similarly.	Since the results showed that the EIKEN common-scale scores are nearly equivalent to TOEFL iBT scores, EIKEN examinations can be used for screening purposes and for language proficiency purposes at advanced levels in Japan.
Brown (2012b)	This is a comprehensive book covering various topics related to rubrics. It presents numerous models, types, analyses, and uses of rubrics at program and classroom levels. Using case studies, it illustrates rubric development processes, the analysis of rubric-based results, as well as rubric-based assessment of reliability and validity.	The book can serve as a valuable addition to a language testing course as it can help teachers and researchers who would like to develop their own rubrics or need to learn or teach about the analysis of rubric-based assessment data.

Kondo-Brown, Brown, et al. (2013)	This book covers various issues related to oral performance tests, oral Japanese language, oral proficiency, placement examinations, assessing written skills, rubric development, ePortfolios, scoring methods for composition tests, teaching and evaluating translation, standards-based final examinations, self-assessment, Japanese cultural testing, assessment for service-learning, and assessment of learner autonomy.	The book will potentially be useful for teachers, teacher trainers, graduate students, and postgraduates. Also, language testers and researchers interested in research in the assessment field can use the book for redesigning and redeveloping oral assessment tests. It could also be used as a coursebook included in the syllabus of language testing course.
Brown (2014d)	This book contains three sections: section one provides an introduction, ways to start research, and gathering, compiling, and coding data; the second section discusses analyzing quantitative, qualitative, and mixed-method data; and section three elaborates on research results, reports, and research dissemination.	The book can be used as a part of instructional materials for a research course, as it is a comprehensive resource book for instruction in such a course and also for self-study for researchers, graduate students, and post-graduates, providing both theoretical and practical bases for research.
Brown (2015d)	The study aims at developing the pilot project and validates the Samoan oral proficiency interview designed to assess the needs of the students at Samoan language center. The Samoan oral proficiency interview is examined in terms of validity and consistency of the reliability of scores.	The study has implications for test developers. They can develop more detailed and specific rubrics for such tests and consider the needs of the learners in other micro-contexts so that more diagnostic and achievement feedback can be prepared for language pedagogy.
Brown & Coombe (2015)	Containing 36 chapters and organized into four main sections, the book provides an exhaustive overview of L2 research methods and covers issues such as doing research, using research, data gathering methods, publishing your research, and research contexts.	The book has the potential to be used as a coursebook and as a part of syllabus in a research course because it provides an effective research knowledge base for graduate students and postgraduates, as well as the scholars more generally.
McKay & Brown (2015, 2016)	Taking a more novel and exact look at the teaching and assessing of English as an international language, the book explains specific principles and strategies for teaching and assessing language skills, proficiency and literacy skills, needs analysis, and challenges faced by English language learners and users around the world in general and in the classroom context in particular.	The books can serve the purposes of educators and graduate students, and for those in pre-and in-service courses on language teaching and assessment, as the authors provide valuable guidance and initiation into the details of English as an international language.
Brown (2016c); Trace et al. (2015)	Brown (2016c) zeros in on needs analysis and ESP, including step-by-step processes for performing a needs analysis in ESP and data collection and interpretation procedures. Developing courses in ESP is also detailed from other perspectives in Trace et al. (2015).	The books provides both theory and practice knowledge base for pre-service and in-service teachers, readers, instructors, and researchers because it provides clear example and helpful exercises of real-world applications.
Kondo-Brown & Brown (2017)	Containing 12 chapters, the book examines issues such as needs analysis, attitude, motivation, identity, instructional preferences, curriculum design, materials development, and assessment procedures. Regarding placement test issues, research conducted by Kondo-Brown and Brown (2000), Brown (2007d), and Brown, Hsu, et al. (2017) are also informative.	The book can serve the purposes language teachers and researchers and be used as a primary text or reference for graduate students and postgraduates in the areas of assessment, pedagogy, and curriculum associated with both bilingual and heritage students.
Lanteigne et al., (2021a)	Containing thirty-seven chapters, this collection examines issues such as descriptive statistics, parametric statistical testing, washback, fairness, and construct-irrelevance, higher-order statistics, test use and consequences, high-stakes assessments, placement, classroom assessments, an evidence-based approach to language testing, constructed-response items, assessing the	The book is a comprehensive sourcebook for educators, language testers, L2 researchers, graduate students, and postgraduates. Due to the comprehensive nature of the book and wide coverage of topics on language testing, the book can be used as a coursebook on

	productive skills, summative and formative test issues, placement tests, reading fluency assessment, and other testing relevant issues.	language testing, as it can also add to the theoretical knowledge base for researchers and testers.
Brown & Crowther (2022)	This book provides a comprehensive account of the theoretical and practical rationale for teaching connected speech and examines issues such as transcribing connected speech, word stress, utterance stress and timing, phoneme variations, simple transitions, as well as dropping, inserting, and changing sounds, and the multiple processes involved in connected speech.	The book has implications for teachers and teacher trainers; it can help them develop their phonological literacy at segmental, phrase, and utterance levels and learn about the rules governing connected speech in spoken North American English.

Throughout nearly all the articles Brown has published, there is a clear link between whatever the topic of the article is, the central thesis of his articles and their application to language pedagogy and research statistics; he argues often that whatever research is carried out must be evaluated in conjunction with our professional experiences. That is why Brown (2008d) said that language testing is too important to be left to language testers, that is, administrators, teachers, examinees, and any other relevant stakeholder groups should be involved. Only through such cooperation, working together, and doing analysis of the entire testing context can we arrive at defensible consequences. In looking back at this entire review, we (Ali Panahi and Hassan Mohebbi) want to emphasize that no such systematic review can be a one-size-fits-all analysis of his work. Due to the nature of his contributions, which are extremely extensive, other scholars reviewing his work might come up with entirely different implications—in addition to ours—for every single individual publication. As a consequence, although we have carefully tried to include everything that should be included, we do not claim that our systematic review is comprehensive. Nor do we claim that the various implications of the review that we discuss will apply and generalize to all teachers, administrator, and researchers in all settings.

Phase II: Discussion and Personal Reflection (James Dean Brown)

Many thoughts ran through my mind as I read through the review of my work above. Among them, I recognized what an enormous task the authors had taken on for themselves and how grateful I am for their dedication to that task. Another idea that occurred to me was that they had artfully classified my work into categories that were mentioned repeatedly. Such classifications largely sidestep any notion of how my publications in each topic area changed and evolved as time went by. From my perspective, I was writing in streams of research that considered different aspects of each of five strands (language testing, criterion-referenced testing, curriculum development, research methods, and connected speech) and covered each topic area from a different perspective or advanced the research step-by-step from article to article. Granted, their **Introduction** section did a stellar job of summarizing the overall history of my work, even delving a bit deeper in one paragraph listing some of my cloze research studies. However, that paragraph on my cloze testing research ended by saying that Brown (2013a), which reviewed my research to that point, concluded in their words "that cloze tests function appropriately as one type of overall ESL proficiency tests." While that statement is largely true, a closer look at the article and indeed at the entire string of my cloze studies will reveal that my views were evolving and were much more complicated. One purpose of this reflection then will be to demonstrate what was going on within all those categories of publications from a personal perspective, or put another way, to show some of the connections between papers that exist at a human level in research and writing.

To provide a frame around what I am talking about, I will step back for a moment and explain how I view the research process. Like *invention*, which has been said to be one percent inspiration and 99 percent perspiration (often attributed to Thomas Edison), *research* to me involves about *inspiration* and *perspiration*, but also requires *revelation*. Formulaically:

Research = Inspiration + Perspiration + Revelation

In brief, *inspiration* comes before or at the beginning of any particular project and motivates ideas for new ways to think about topics or new questions to answer by conducting research. *Perspiration*, of course, represents the huge amount of hard work necessary to carry out research and write papers and books. And finally, *revelation* is what emerges or reveals itself in the process of working on one project that leads to ideas or questions for other projects in a constant ongoing manner. In a sense, revelation is inspiration, but it is the kind of inspiration revealed in the process of doing one project that leads to others.

Inspiration

As mentioned above, inspiration is anything that helps the researcher before or at the beginning of a project to see new ways of thinking about topics within topic areas or questions that can be answered through research. Inspiration can come from many places and often occurs at odd and unexpected moments. That said, the probability of finding inspiration increases (a) if you spend time with people in your profession and listen carefully to what they have to say, (b) if you keep abreast of and pay attention to the latest literature related to your topics of interest, and (c) if you carefully observe what is happening around you in your work. I will abbreviate those three sources of inspiration as *people*, *papers*, and *processes*.

In this section, I will explain how people, papers, and processes inspired me in six of the topics shown in Table 1 above: (a) language testing and assessment; (b) research and statistics; (c) curriculum development and language program development; (d) cloze tests; (e) connected speech and reduced forms; and (f) pragmatics tests and issues. The other 17 topic areas in Table 1 seem to me to be subcategories of these six in terms of how I viewed them in my career. Let's consider these six topic areas in more detail to see how I was inspired to get involved with each.

Language testing and assessment. On the last day of my first ESL teaching methods course at UCLA, Professor Russ Campbell was talking about things you can do with training in teaching ESL. He talked of course about teaching, but also about materials development, administering programs, doing teacher training, etc. [Let me step back a second to point out that I was a French Horn player for most of my life up to this point, even attending the Oberlin Conservatory for two years and playing professionally in US army bands. As a French horn player, I had noticed that people who chose unpopular instruments in the orchestra like French horn, oboe, bassoon, and viola, were always chosen in any selection process, while selection for popular instruments like violin, flute, trumpet, etc. was much more ruthless. As a result, I was looking for the French horn of the ESL field (i.e., an unpopular specialization) as Russ was talking.] Toward the end of his lecture, he said something that really made my ears perk up, "Oh yeah, and there is one other thing that you can do in the field, but most people in language teaching are not interested because it involves a lot of math, yet every language program or department needs one. That is *language testing*." Since I had always liked math and was pretty good at it, I realized that language testing just might be my ESL French horn, and I was off and running. Inspiration!

During the second term in my MA program at the University of California at Los Angeles (UCLA), I took a language testing (LT) course that was so poorly taught that I had to get myself a mastery book on statistics and set up a study group with the other students so that we could learn the material and pass the course. That poorly taught course inspired me too (in a negative way) to migrate over to the Education Department, where I took a well-taught course in testing offered by W. James Popham. Since he was one of the fathers of criterion-referenced testing (CRT) (see Popham & Husek, 1969), it is easy to see who inspired my interest in CRT. Popham was excellent on the practical and conceptual aspects of CRT, but it wasn't until I took an advanced testing course with Richard Shavelson (who had graduated from Stanford and studied with Lee J. Cronbach) that I learned about Cronbach's G theory and its sophisticated theoretical and mathematical connections to CRT and norm-refenced testing (NRT). Once I was in the PhD program at UCLA, it turned out that my tennis partner John Dermody (a former student in the MA program) had become the editor for English Language Services (ELS) publications, and one way or another, I was subcontracted to develop placement and achievement tests for two ELS book series (each involving six course books): *The New English Course* and *Welcome to English Course*. One day, talking to Earl Rand about these projects, he suggested that I use Rasch analysis for developing the NRT placement tests. Since Shavelson had also taught me about Item Response Theory (IRT), I knew the basics of doing such analyses and another sub-strand of my research was inspired that showed up repeatedly in many of my studies. Step by step, you can see how I was inspired in various ways by Campbell, Popham, Shavelson, Dermody, and Rand to become a language tester who knew the theory and practice of

psychometrics, CRT, G theory, and IRT. Thus people, papers, and processes inspired me and prepared me to do *language testing and assessment*.

Research and statistics. As mentioned above, after having had to learn basic statistics from a mastery book, I ended up turning to the Education Department at UCLA to take three levels of testing courses, three levels of basic research design and statistics courses, and then advanced courses in ANOVA, Regression, Survey Research, etc. During that whole process, I had the pleasure (because he was such an inspiring and straightforward explainer of research design and statistics) of taking five courses with Richard Shavelson and two from his colleague Noreen Webb. Those two professors instilled in me the very conservative (even skeptical) attitudes toward statistics that are peppered throughout my papers and books on research methods. Thus, people and processes inspired me to explain *research design and statistics* in straightforward terms that would be useful to language professionals.

Curriculum development and language program development. During my four years of coursework at UCLA, I managed to feed my wife and my family by working as an instructor of ESL at Marymount Palos Verdes College (MPVC). At MPVC, all teaching staff were required to attend regular workshops that MPVC provided. One such workshop (presented by someone whose name I cannot remember) explained in depth how to write course objectives; that in turn inspired me to read Mager (1962) on writing educational objectives and Popham's (1975, 1978) books on program evaluation and CRT. Those readings eventually lead me, as one of two Senior Scholars at the Guangzhou English Language Program (GELC was my first job after finishing my coursework at UCLA), to run workshops on curriculum development for my colleagues. Those workshops served as the basis for developing the curriculum for our 15 English-for- science-and-technology courses at GELC (each including needs analysis, objectives, testing, materials, and evaluation). That along with experiences running an MA program for Florida State University in Saudi Arabia (and observing curriculum development efforts there at various levels) and serving as Director (coordinating all curriculum development) of the eight English- for-academic-purposes courses in the English Language Institute (ELI) at the University of Hawai‘i at Mānoa (UHM) inspired me to write my second book (Brown, 1995g) and a number of articles on curriculum development. Thus, my experiences and observing myself and my colleagues struggling through various curriculum development processes in the real world inspired me to research and write about *language curriculum*.

Cloze tests. After meeting John Oller during a summer session at UCLA, I read some of his early research on cloze testing (Oller, 1973; Oller & Conrad, 1971), and that raised my awareness of cloze testing as a line of research. I had also noticed that data gathering was a very difficult and time-consuming part of much language research. Quite honestly, this observation along with my reading led me to choose the cloze testing topic for my MA thesis (see below to find out how this turned out) at least partly because cloze tests are relatively easy to develop, administer, and score, all of which made data gathering fairly easy. Thus, people, papers, and processes inspired me to do *cloze testing research*.

Connected speech and reduced forms. J. Donald Bowen was one of my early mentors and chaired my MA thesis committee. He sparked my interest in and helped me understand a problem I had noticed in the ESL classes I was teaching at MPVC. In conversations he mentioned something he covered in depth in his pronunciation book (Bowen, 1975) that he

called *informal speech* and what he also called *reduction*. A few years later while I was teaching in China, one student in my speaking course asked, "Why can I understand you when you talk to the class, but not understand when you talk to the other Americans?" The combination of what Bowen had revealed to me and that student's question led me to collect (with Ann Hilferty) ideas from our own experience and from colleagues and to produce a list of *reduced forms* (see Brown & Crowther, 2022, p. xiii, for that original list). We started to teach those reduced forms in our speaking courses, then did a study (Brown & Hilferty, 1982) that led to a chain of research that I did throughout my career, including three books (Brown, 2012a; Brown & Crowther, 2022; Brown & Kondo-Brown, 2006).[1] Thus, people, papers, and processes inspired me to work in the area of *connected speech* (a more accurate name for what we originally called *reduced forms*).

Pragmatics tests and issues. Lyle Bachman (1990, p. 89) divided communicative language ability into three subcategories: strategic competence, psychophysiological mechanisms, and language competence. Language competence was further subdivided into organizational competence and pragmatic competence. Based on that, Thom Hudson and I were inspired to actually try to systematically test pragmatic competence. Conveniently, we could turn to our colleague, Gabi Kasper, in the Department of Second Language Studies (SLS) at UHM, who was a well-known expert in the area of pragmatics, and she supplied us with more than enough reading material for us to realize that research in pragmatics was limited in the sophistication of their tools for measuring pragmatic competence—largely relying on discourse completion tasks. As a result, we developed six different measures of pragmatic competence and then did research on the reliability and validity of these measures. This further inspired a string of studies on testing pragmatic competence produced by us, our students, and others. It all began with the realization that one of Bachman's categories was seldom tested and that one of the leading experts on pragmatics had an office three doors down from mine. Thus, people and papers inspired Thom and I to work on *pragmatics testing*.

Inspiration coda. In this section, I tried to show how people, papers, and processes, in various combinations, *inspired* me in six of the topic areas listed in Table 1. I hope I managed to make clear how inspiration can come from multiple sources, but I also want to stress that inspiration can come in many sizes ranging from small comments that resulted in big consequences to big inspirations that lead to a number of small consequences. An example of the former is the simple question mentioned above from a Chinese student that resulted in an entire strand of research on connected speech. An example of the latter is the large impact that the five courses I took with Richard Shavelson had on my statistical philosophy and knowledge, an impact that shows up in many ways in many places in my work throughout the years.

Perspiration

As mentioned above, *perspiration* involves putting in the work required for doing research, as well as writing papers and books. Probably because I flunked out of my undergraduate degree after two years in the Oberlin Conservatory, five years later when I returned to college, I was a very motivated undergraduate and then graduate student, and I learned in that process how to

[1] See Brown (2023) discussion of a chain of 12 publications that I found inspiring and even essential to my understanding of connected speech and how those publications influenced my research over the years.

work hard on projects like my course projects, MA thesis, and PhD dissertation. Once I became a professor, I recalled that I had witnessed several of my favorite young professors at UCLA lose their jobs for lack of publications. These young professors were not lazy people. After all, they had worked very hard to get PhD degrees from top notch universities. Nonetheless, the *publish or perish* dictum had seen them perish. I concluded from having watched them in the department that they simply had trouble managing their time and getting themselves organized to sit down and do the research and writing part of their job.

Rules that helped me get my research and writing job done. Once I became a professor myself, I was in constant fear that I might *flunk out* again, so I set myself some rules that I knew from working hard during my graduate studies would help me get organized to sit down and do the research and writing part of my job:

1. Set aside time *every single day* for research and writing. Even if you are not eager to work, sit down, open your computer, and at least read through what you have so far, or jot down some notes, or open up your data and have a look—every single day. I did this for nearly four decades, and it worked for me.
2. Work on more than one project at a time. Most research and writing projects are necessarily long and drawn out—taking months or even years. If you wait to finish one project before beginning another, you will not be very productive. I preferred to have three to five projects running at all times. When one would finish, I would start up another.
3. Having three to five projects moving along simultaneously also helped to create variety in my workday. I found it very useful to work on different aspects of various projects throughout my working hours. For example, I might start by gathering and jotting down ideas for one project for a few minutes, then shift to writing a chunk of another project, and then do some proofreading on another project and end my workday with some data entry or analysis in Excel or SPSS (statistical analysis software). Shifting through different projects and different types of work helped keep my energy and interest levels up as all of my projects moved along incrementally.
4. One other side benefit of working on smaller bites of multiple projects each day was that all of my projects were perking away in my brain even while I was not working directly on them, such that I often found ideas popping into my head related to one project or another while I was doing other things. Thus, I needed note pads scattered around everywhere at work and at home so I could jot down these bits and pieces. Since all the materials for each project had its own slot above my desk at work. I could easily sort the stack of notes that accumulated in my pockets into the appropriate slots for later reference.
5. I also found it helpful to take physical breaks for 5 minutes about once an hour and move around to get my blood circulating. To do so, I would get up and go to the break room to get coffee or walk out to the trash bin to empty my office trashcan, or just do a couple of dozen pushups right there in my office.
6. In addition, if you are not enjoying your work, you're not doing it right. Try something new to spice it up. For example, I once found myself in the doldrums dreading even sitting down to work, so I got a small stereo and a set of headphones and started listening to whatever *instrumental* music suited my mood: sometimes slow baroque music worked, other times I needed hard-driving music. The point is that I changed things up and got back into enjoying

my work. Incidentally, the headphones also blocked out the outside world, which was often helpful.

7. When duty calls and you must take your turn at administrative duties, don't be afraid to protect your research time. Even during the ten years when I had extensive administrative duties (as Director of the ELI, Chairman of the SLS Department, Director of the NFLRC), I arranged to have certain times of every day when my office door was closed and I was not available (usually mornings with the department secretaries running interference except for dire emergencies). However, it is equally important to have times when you are regularly 100% available to students, colleagues, and above all to the secretaries. After sitting alone during the whole morning with my door closed, I was usually happy to talk to people, have appointments, hold meetings, do mindless paperwork, and of course teach my classes. It helped that my classes were always scheduled late in the afternoon because I always found that I could wring out a few ounces of energy my teaching.

Coda for perspiration. Naturally, all of this became easier once I was a professor and was paid to do research; it became even easier once I was a full professor and had the seniority to arrange and control my working and teaching schedules. But even when I was a young insecure graduate student, instructor, and assistant professor with a family, I would wake up early at 5:00 am, and work for a couple of hours in the pattern described earlier, and then wake up the kids and take them to school. The important thing to note is that throughout my career, because it was obvious to me that *publish or perish* was real, I organized myself to put in the time to work on research and writing every single day, always moving ahead on multiple projects a bit at a time. For more depth on the ideas discussed in this section and other related topics, see Brown (2014d, pp. 205-236).

Revelation

Revelation is the driver that leads to new knowledge from paper to paper always building on what came before by summarizing, clarifying, correcting, expanding, adjusting, combining, elaborating, and exemplifying—especially in examining the similarities and differences between and within studies. Revelation requires that you: let the data talk to you so that you don't get stuck in preconceived ideas; learn from mistakes so you don't repeat them; listen carefully to students and colleagues; pay attention to what your mind gives you when you are not working; do research collaboratively with others; talk about your research at conferences or elsewhere and pay attention to how people respond and ask questions; be ready to do further follow-up research; and encourage others to run with any research ideas your studies may have inspired.

How revelations in each study connect to those that follow. In Brown (2002f), I started to reflect on how my cloze research studies were all linked head-to-tail with each other over the previous years. I started out in the 1970s wondering how something as simple as a cloze procedure could provide a relatively sound measure of overall English language proficiency. It all began with my MA thesis, which was summarized and published in Brown (1980).

Brown (1980) compared four methods for scoring cloze tests (exact-answer, acceptable-answer, clozentropy, and multiple-choice) and concluded that the exact-answer scoring method was probably the best overall based on a number of test characteristics (usability, item discrimination, item facility, reliability, standard error of measurement, & criterion-related

validity). I later realized that the study was fundamentally flawed because I had overlooked one very important variable: passage difficulty, which would crucially determine how the scores would be distributed and in turn affect the relative values of my descriptive, item analysis, reliability, and validity statistics for the four scoring methods. Thus, if my passage had been easier or more difficult, a different scoring method would likely have appeared to be best.

In Brown (1983a/1984a), my mistake of not considering passage difficulty led me to administer that same cloze test to different groups of students with substantially different ranges of ability to see how that would affect descriptive statistics, reliability, and criterion-related validity. The results showed clearly that the same cloze test administered to groups with varying ranges of ability would sometimes result in very high levels of reliability and validity and other times in very low levels, depending on the range of abilities in the group, as measured by the standard deviation and range. More generally, I realized that the degree to which a sample of items fits the proficiency levels of the students is crucial to what happens to the descriptive statistics, reliability, and validity coefficients.

In Brown (1983b), I used my 1980 data to do two studies (reported in one paper for editorial reasons), in which I examined the relationship of different aspects of linguistic cohesion in the items to the scoring methods as well as the effects of the scoring methods on a wide variety of reliability estimates. Because I still was not grappling with passage difficulty, I later realized that these two studies were as flawed as the first. However, I did notice one useful thing: the K-R21 estimate consistently and substantially underestimated the reliability of cloze tests as compared to all other estimates of reliability that I had calculated.

To help me understand these aberrant reliability results, I turned to the original Kuder and Richardson (1937) article, where I learned that one fundamental difference between K-R20 and 21 was that K-R21 assumed that items must be of equal difficulty while other formulas did not. Thus, K-R21 could reasonably be expected to provide good estimates of reliability for multiple-choice (MC) tests because such an assumption would be tenable because we create MC tests by pretesting and selecting items with item facility values (IF) ranging *narrowly* (e.g., from say .30 to .70, or 30% answering correctly to 70% answering correctly). However, the equal item variances assumption is not tenable for cloze tests because the items range *widely* from very difficult (IF = .00, i.e., nobody answering correctly) to very easy (IF = 1.00, i.e., everybody answering correctly). Thus, I realized that these serious underestimates of K-R21 could be accounted for by the fact that many cloze items violate the equal difficulty assumption (later explained in Brown, 1993d).

I had also claimed in Brown (1983b) that the blanks in cloze tests provide a reasonably representative sample of the linguistic material in the passages—regardless of the starting point for the deleted words. However, given my new understanding that some items on cloze were doing nothing (in terms of spreading students out) because nobody was answering them correctly (IF=.00) and others were doing nothing because everyone was answering them correctly (IF=1.00), I had to admit to myself that, regardless of what the items appeared be testing linguistically, since many items might not be functioning at all in terms of test variance (or spreading students out), those items that were functioning well might not be representative of the linguistic material in the original passage. Put another way, I had realized that, if only some of the items on a cloze test are functioning well for a particular group of students, the

variance produced by those items, and the variance on the cloze test as a whole, might only be coming from those few items that are functioning well. Thus Brown (1983b) led me to realize that selecting different samples of items, even from the same passage, could result in cloze tests that behaved quite differently.

I then turned my full attention to the importance of item analysis in cloze testing, which resulted in Brown (1988a) where I systematically used administrations of five different 50 item deletion patterns from the same passage to select, from among the 250 items, those items that discriminated well (or spread students out as measured by item discrimination) to create a sixth "well-tailored" 50-item cloze passage. When I then administered that tailored cloze, I found that it was far more reliable and valid than any of the five earlier versions.

Also, based on what I had learned in Brown (1983b), I tried to understand in Brown (1989b) how item discrimination worked at the linguistic level. Since item discrimination is related to item difficulty (i.e., items that 50% of student answer tend to discriminate better than very difficult or very easy items that nobody or everybody answers correctly), I used regression analysis to examine the relationships between the linguistics characteristic of 150 cloze items (from five 30-item cloze tests administered to 179 Japanese university students) and item difficulty. The results were interesting but not very powerful, that is, the correlations between individual linguistic variables and item difficulty were only .14 to .51, and the multiple regression analysis showed that, at best, four of the linguistic variables could predict only 32 percent of the variance in item difficulties.

Thus, I came to wonder if understanding the item level in cloze could provide only part of the picture, and in Brown (1993d), I turned to the whole passage level by marshalling the cooperation of a number of colleagues who were willing to administer 50 30-item cloze tests to 2298 randomly assigned students from 18 universities across Japan. In that study, I looked at how 50 randomly selected passages (from a US public library) developed into cloze tests would naturally vary in terms of statistical characteristics (e.g., passage difficulty, reliability, etc.). The Brown (1989b) study also led me to wonder what would happen if I studied multiple passages with different difficulty levels simultaneously administered to groups of students at different proficiency levels, which occurred in Brown, Yamashiro, et al. (1999, 2001) and Brown (2002f, 2013a).[2]

In Brown (1998c), I again studied the data I had used in Brown (1993d), but this time, unlike Brown (1989b), I analyzed the effects of linguistic variables including readability indices on passage difficulty. Unlike the 1989b study, the individual correlation coefficients among variables were much higher, and the multiple regression analysis with four linguistic independent variables accounting for 55% of the variance in the dependent variable, passage difficulty.

That study led me, in turn, to wonder what differences might exist between students from very different language groups, which set me to studying the relationships between linguistic variables including traditional readability indices (e.g., the Fry scale) and passage difficulty for Japanese university students in Brown (1992g), as well as for Russian students in Brown,

[2] See Brown 2013a for a more detailed discussion to the chain of papers up to that date of publication.

Janssen, et al. (2019), and for both Japanese and Russian students in Trace, Brown, et al. (2017).

I also stepped back and pondered the whole sweep of my cloze research in Brown (2002f, 2013a) while focusing on the effects of non-functioning items, or what I called *turned off items*, on the distributions of scores and reliability of the 50 cloze tests I had administered in earlier studies.

Step by step, I have shown how my cloze studies connected, but one other worry that kept nagging at me during this whole process was the fact that I knew researchers, especially in second language acquisition, had been using my original 1980 cloze test as a reliable and valid measure of overall English language proficiency (citing my 1980 study to bolster that contention). The reason I was worried was that my whole string of research had clearly demonstrated to me that much depended on the relationship between passage difficulty and the proficiency level and range of abilities of the examinees. I was able to pull all of this together by working with Theres Grüter on what is probably my last cloze study ever (Brown & Grüter, 2022), which examines data from a wide variety of settings (EFL and ESL) from widely differing proficiency levels and ranges. Based on 1724 examinees in 19 data sets gathered from 1977 to 2015. This study corrects the misconception that my 1980 cloze test was a reliable and valid measure of overall proficiency, in and of itself. The paper shows instead how that cloze test operates in different contexts and provides researchers with the tools to put their results within the context of all the available data.

Revelation coda. In this section, I have shown how my cloze research studies flowed head-to-tail from one study to the next, and sometimes from one study to a number of others. While all of the studies in this section can correctly be said to be about cloze testing (and perhaps about their reliability and validity), I hope that you can now see that there have been revelations in each that led to those studies that followed and that, over time, the totality of the studies is much greater than the sum of its parts. In other words, to conclude, as my co-authors did that I found "that cloze tests function appropriately as one type of overall ESL proficiency tests" in Brown (2013a), while true, is necessarily very reductive. Adding the phrase *depending on how well the items fit the proficiency levels and ranges of the examinees involved* would make it much more accurate. I hope that is clear now.

I am sure I could show similar head-to-tail connections in all of the research strands that I have worked on. For example, Brown (2023) explains how 12 primary references (and others) influenced my development of the connected-speech strand of my work—again showing a steady stream of work that produced revelations that led to further work.

On a related note, one researcher's revelations can clearly serve as another researcher's inspiration. Just out of curiosity, I just turned to Google Scholar to find out how many people have cited my cloze articles. My first 1980 article has been cited as of today in 282 articles, while all of my articles with the word *cloze* in the title have been cited a total of 1022 times. I hope this represents at least some inspirations from my articles leading other researchers to make their own connections and revelations.

Conclusion

As always when I am writing, I have a particular audience in my mind—the people I am addressing. As you may have guessed by now, that audience, in this case, was novice or mid-

career researchers who might have stumbled into this article and benefit from my reflections. Without starting out with the intention of doing so, this reflection focused on the human side of doing research and writing. I was inspired to break it down into inspiration, perspiration, and revelation by the Thomas Edison saying. To clarify how those three concepts functioned during my research career, I ended up providing: (a) a list of three ways to improve the probability of coming up with inspiration for research questions or new perspectives on your topic areas of interest; (b) an explanation of how six of my research strands were inspired by people, papers, and processes; (c) a list of seven rules that helped me to publish rather than perish; and (d) a discussion of some of the ways each of my cloze research projects revealed unexpected knowledge that connected directly to my other studies and the studies of people I have influenced.

I don't want to leave the impression that I did all of this completely alone. True, being a language tester was very lonely when I first started out, after all ever ESL program or department needed *one*, and only one. But once I discovered and began attending the yearly Language Testing Research Colloquium, I discovered a whole community of like-minded individuals who were spread out around the world. The advent of *Language Testing Journal* and later the *Language Assessment Quarterly*, helped to legitimize what we had all been doing all along. Soon local language testing organizations were sprouting up in various regions and in specific countries. For example, I was intimately involved in the founding of the Language Testing and Evaluation NSig within the Japan Association on Language Teaching (JALT), which was founded immediately after I presented a plenary speech at the JALT annual conference (on problems with the university entrance exam system in Japan). I also have a sneaking suspicion that the Japan Language Testing association founded shortly thereafter was founded at least in part in reaction to the new JALT NSig. Thus, I have seen language testing, and indeed all of my areas of interest, grow in size and stature as sub-fields within the broader fields of Second Language Studies and Applied Linguistics. I have been proud to belong to such a vibrant field and hope that I have contributed in some small way to its growth. I also hope that some of the ideas I have presented in this reflection will prove useful to those members of my intended audience who have read this far. *Best of luck with your researching!*

ORCID

https://orcid.org/0000-0003-2211-241X

https://orcid.org/0000-0001-7167-0026

https://orcid.org/0000-0003-3661-1690

Acknowledgements

Not applicable.

Funding

Not applicable.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

Rights and Permissions

References

Alderson, J. C. (1978). *A study of the cloze procedure with native and nonnative speakers of English*. Unpublished Ph.D. thesis, University of Edinburgh.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford.

Bailey, K.M, & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp.236-245). Multilingual Matters.

Bloom, B. (Ed.). (1956). *Taxonomy of educational objectives: Handbook 1. Cognitive domain*. Longman**.**

Bowen, J. D. (1975). Patterns of English pronunciation. Newbury House.

Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal, 64*(3), 311-317. https://doi.org/10.2307/324497

Brown, J. D. (1983a, March 15-20). *A cloze is a cloze is a cloze?* [Paper presentation]. Proceedings of the annual convention of teachers of English to speakers of other languages, Toronto, Canada.

Brown, J.D. (1983b). A closer look at cloze. In J. W. Oller (Ed.), *Issues in language testing* (pp. 327-350). Newbury House.

Brown, J.D. (1983c). A norm-referenced engineering reading test. In A.K. Pugh & J. M. Ulijn (Eds.), *Reading for Professional Purposes: studies in native and foreign languages* (pp.213-222). Heinemann Educational.

Brown, J. D. (1984a). A cloze is a cloze is a cloze? In J. Handscombe, R. A. Orem, & B. P. Taylor (Eds.), *On TESOL '83* (pp. 109-113). TESOL.

Brown, J.D. (1984b). Criterion-referenced language tests: What, how and why? *Gulf Area TESOL Bi-annual, I,* 32- 34.

Brown, J. D. (1988a). Tailored cloze: Improved with classical item analysis techniques. *Language Testing, 5*(1), 19-31. https://doi.org/10.1177/026553228800500102

Brown, J. D. (1988b).Components of engineering-English reading ability? *System, 16*(2), 193-200. https://doi.org/10.1016/0346-251X(88)90033-4

Brown, J. D. (1988c). *Understanding research in second language learning: A teacher's guide to statistics and research design.* Cambridge University Press.

Brown, J. D. (1988d). What makes a cloze item difficult? *University of Hawaii working papers in ESL, 7*(2), 17-39.

Brown, J. D. (1989a). Language program evaluation: A synthesis of existing possibilities. In R.K. Johnson (Ed.), *The second language curriculum* (pp.222-241). Cambridge University Press. https://doi.org/10.1017/CBO9781139524520.016

Brown, J. D. (1989b).Cloze item difficulty. *JALT Journal, 11*(1), 46-67.

Brown, J. D. (1989c).Rating writing samples: An interdepartmental perspective. *University of Hawai'i Working Papers in ESL, 8*(2), 119-141.

Brown, J. D. (1989d). Improving ESL placement tests using two perspectives. *TESOL Quarterly, 23*(1), 65-83. https://doi.org/10.2307/3587508

Brown, J. D. (1989e). Criterion-referenced test reliability. *University of Hawai'i Working Papers in English as a Second Language, 8* (1), 81-115.

Brown, J. D. (1989f). A systematically designed curriculum for introductory academic listening. In C. Donovan & C. Serizawa (Eds.), *Program development in ESL: ESL textbooks for summer sessions* (pp.1-68). University of Hawaii.

Brown, J. D. (1989h). 1989 Manoa writing placement examination. *Manoa Writing Board Technical Report, 5*.

Brown, J. D. (1990a). Placement of ESL students in writing-across-the-curriculum programs. *University of Hawai'i working Papers in ESL, 9* (1), 35-58.

Brown, J. D. (1990b). The use of multiple t-tests in language research. *TESOL Quarterly, 24*(4), 770-773. https://doi.org/10.2307/3587135

Brown, J. D. (1990c). Short-cut estimators of criterion-referenced test consistency. *Language Testing, 7*(1), 77-97. https://doi.org/10.1177/026553229000700101O

Brown, J. D. (1990d). Where do tests fit into language programs? *JALT Journal, 12*(1), 121-140.

Brown, J. D. (1990e). Placement Test (ELIPT): subtests for ESL listening, reading and writing. *University of Hawaii Working Papers in ESL, 9*.

Brown, J.D. (1991a).Performance of ESL students on a state minimal competency test. *Issues in Applied Linguistics, 2*(1). https://doi.org/10.5070/L421005131

Brown, J. D. (1991b). Do English and ESL faculties rate writing samples differently? *TESOL Quarterly, 25*(4), 587-603. https://doi.org/10.2307/3587078

Brown, J. D. (1991c). Statistics as a foreign language: Part 1: What to look for in reading statistical language studies. *TESOL Quarterly, 25*(4), 569-586. https://doi.org/10.2307/3587077

Brown, J. D. (1991d). A comprehensive criterion-referenced language testing project. *University of Hawaii Working Papers in ESL, 10*(1), 95-125.

Brown, J. D. (1992a). Using computers in language testing. *Cross Currents, 19*(1), 92-99.

Brown, J. D. (1992b). Classroom-centered language testing. *TESOL journal, 1*(3), 12-15.

Brown, J. D. (1992c).Technology and language education in the twenty-first century: Media, message, and method. *Language Laboratory, 29*, 1-22.

Brown, J. D. (1992d).What is research. *TESOL Matters, 2*(5),1-10.

Brown, J. (1992e). The biggest problems TESOL members see facing ESL/EFL teachers today. *TESOL Matters, 2*(2), 1-5.

Brown, J. D. (1992f). Identifying ESL students at linguistic risk on a state minimal competency test. *TESOL Quarterly, 26*(1), 167–172. https://doi.org/10.2307/3587385

Brown, J.D. (1992g, March 1). *What text characteristics predict human performance on cloze test items* [Paper presentation]. Proceedings of the conference on second language research in Japan.

Brown, J. D. (1992h).Statistics as a foreign language—Part 2: More things to consider in reading statistical language studies. *TESOL Quarterly, 26*(4), 629-664. https://doi.org/10.2307/3586867

Brown, J. D. (1993a). The social meaning in language curriculum, of language curriculum, and through language curriculum. In J. E. Alatis (Ed.), *Georgetown University round table on languages and linguistics: language, communication, and social meaning* (pp.117-134). Georgetown University Press.

Brown, J. D. (1993b). Language test hysteria in Japan. *The Language Teacher, 17*(12), 41-43.

Brown, J. D. (1993c). *A Comprehensive Criterion-Referenced* [paper presentation]. Proceedings of Papers from the 1990 Language Testing Research Colloquium: Dedicated in Memory of Michael Canale.

Brown, J. D. (1993d). What are the characteristics of natural cloze tests? *Language Testing, 10*(2), 93–116. https://doi.org/10.1177/026553229301000201

Brown, J.D., (1993e). A comprehensive criterion-referenced language testing project. In D. Douglas, & C. Chapelle (Eds.), *A new decade of language testing research* (pp.163-184). TESOL.

Brown, J. D. (1995a). *The element of language curriculum: A systematic approach to program development.* Heinle & Heinle.

Brown, J. D. (1995b). *Differences between norm-referenced and criterion-referenced tests* [Paper presentation]. Proceedings of the 21st JALT international conference on language learning/teaching, Tokyo, Japan.

Brown, J. D. (1995c). *Developing norm-referenced tests for program-level decision making* [Paper presentation]. Proceedings of the 21st JALT international conference on language learning/teaching, Tokyo, Japan.

Brown, J. D. (1995d). English language entrance examinations at Japanese universities: Interview with James Dean Brown. *The Language Teacher, 19*(11), 7-9.

Brown, J.D. (1995e). Understanding and developing language tests. *Studies in Second Language Acquisition, 17*(1), 103-104. https://doi.org/10.1017/S0272263100013899

Brown, J.D. (1995f). Language program evaluation: Decisions, problems and solutions. *Annual Review of Applied Linguistics, 15,* 227-248. https://doi.org/10.1017/S0267190500002701

Brown, J. D. (1995g). *The Elements of Language Curriculum Development*. International Thomson Publishing Company.

Brown, J. D. (1995h). A gaijin teacher's guide to the vocabulary of entrance exams. *Language Teachers, Kyoto,- JALT, 19*, 25-26.

Brown, J. D. (1995i). The TESOL Quirks project. *University of Hawaii Working Papers in TESOL, 13*(2), 121-125.

Brown, J.D. (1995j). Differences between norm-referenced and criterion-referenced tests? In J.D. Brown & S.O, Yamashita (Eds.). *Language Testing in Japan* (pp. 12-19). Tokyo, Japan Association for Language Teaching.

Brown, J. D. (1996a). *Testing in language programs.* Prentice Hall Regents.

Brown, J. D. (1996b). Japanese entrance exams: A measurement problem. *Daily Yomiuri*, 15.

Brown, J. D. (1996c). Fluency development. *JALT, 95*, 174-179.

Brown, J. D. (1996d -November 22).*English language entrance examinations in Japan: problems and solutions*. Proceedings of the JALT international conference on language teaching/learning, Nagoya, Japan.

Brown, J. D. (1997a). Designing a language study. *Classroom Teachers and Classroom Research, 25*, 55-70.

Brown, J. D. (1997b). Statistics corner: Questions and answers about language testing statistics: Reliability of surveys. *Shiken: JALT Testing & Evaluation SIG Newsletter, 1*(2), 18-21.

Brown, J.D. (1997c). Designing surveys for language programs. *Classroom Teachers and Classroom Research, 24.*

Brown, J. D. (1997d).Testing washback in language education. *PASAA Journal, 27,* 64-79.
Brown, J. D. (1997e). Statistics corner: Questions and answers about language testing statistics: Skewness and kurtosis. *Shiken: JALT Testing & Evaluation SIG Newsletter, 1*(1), 20-23.
Brown, J. D. (1997f). Do tests washback on the language classroom. *The TESOLANZ Journal, 5*(5), 63-80.
Brown, J. D. (1997g). The washback effect of language tests. *University of Hawaii working papers in ESL, 16*(1), 27-45.
Brown, J. D. (1997h). Computers in language testing: Present research and some future directions. *Language Learning and Technology, 1*(1), 44-59.
Brown, J. D. (1998a). *New ways of classroom assessment: New ways in TESOL series II. Innovative classroom techniques*. *TESOL.*
Brown, J. D. (1998b). Statistics corner: Questions and answers about language testing statistics: pragmatics Tests and Optimum Test Length. *Shiken, 2*(2), 19-22.
Brown, J. D. (1998c). An EFL readability index. *JALT Journal, 20*(2), 7-36.
Brown, J. D. (1999a). The relative importance of persons, items, subtests and languages to TOEFL test variance. *Language Testing, 16*(2), 217–238. https://doi.org/10.1177/026553229901600205.
Brown, J.D. (1999b). The roles and responsibilities of assessment in foreign language education. *JLTA Journal, 2,* 1-21. https://doi.org/10.20622/jlta.2.0_1
Brown, J. D. (1999c). Statistics corner: Questions and answers about language testing statistics: Standard error vs. standard error of measurement. *JALT Testing & Evaluation SIG Newsletter, 3*(1), 20–25.
Brown, J. D. (2000a). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter, 3*(2), 2–7.
Brown, J. D. (2000b). Statistics corner: Questions and answers about language testing statistics: What issues affect Likert-scale questionnaire formats. *Shiken: JALT Testing & Evaluation SIG Newsletter, 4*(1).
Brown, J. D. (2000c). Statistics corner: Questions and answers about language testing statistics: What is construct validity? *Shiken: JALT Testing & Evaluation SIG Newsletter, 4*(2),8–12.
Brown, J. D. (2000d). Observing pragmatics: Testing and data gathering techniques. *Pragmatics Matters, 1*(2), 5-6.
Brown, J. D. (2000e). Using testing to unify language curriculum. In M. Harvey & S. Kanburoğlu (Eds.), *Achieving a coherent curriculum: Key elements, methods and principles* (pp.220–239). Bilkent University.
Brown, J. D. (2000f). Give second chances in education. *The Daily Yomiuri, 17746 (6).*
Brown, J. D. (2001a). Pragmatics tests: Different purposes, different tests. In K. Rose & G. Kasper (Eds.), *Pragmatics in Language Teaching* (pp. 301–326). Cambridge University Press. https://doi.org/10.1017/CBO9781139524797.020
Brown, J. D. (2001b). *Developing Korean language performance assessments*. University of Hawaii.
Brown, J. D. (2001c). Statistics corner: Questions and answers about language testing statistics: Point-biserial correlation coefficients. *Shiken, 5*(3), 12-15.
Brown, J. D. (2001d). Statistics corner: Questions and answers about language testing statistics: Can we use the Spearman-Brown prophecy formula to defend low reliability? *Shiken, 4*(3),7-11.
Brown, J. D. (2001e). *Using Surveys in Language Programs*. Cambridge University Press.
Brown, J. D. (2001f). Statistics corner: Questions and answers about language testing statistics: What is an eigenvalue? *JALT Testing & Evaluation SIG Newsletter, 5*(1).
Brown, J. D. (2001g).Six types of pragmatics tests in two different contexts. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp.301-325). Cambridge University.
Brown, J. D. (2001h). Statistics corner: Questions and answers about language testing statistics: What is two-stage testing? *JALT Testing & Evaluation SIG Newsletter, 5(2), 8-10.*
Brown, J. D. (2001i). Developing and revising criterion-referenced achievement tests for a textbook series. In T. Hudson & J. D. Brown (Eds.). *A focus on language test development: Expanding the language proficiency construct across a variety of tests*. University of Hawai'i Press.
Brown, J. D. (2002a). *An investigation of second language task-based performance assessments (No. 24).* University of Hawaii Press.
Brown, J. D. (2002b). Statistics corner: Questions and answers about language testing statistics: Extraneous variables and the washback effect. *Shiken, 6*(2), 12-15.
Brown, J. D. (2002c). Statistics corner: Questions and answers about language testing statistics: Distractor efficiency analysis on a spreadsheet. *Shiken, 6*(3), 20-23.
Brown, J. D. (2002d). *Language testing and curriculum development: Purposes, options, effects, and constraints as seen by teachers and administrators* [Paper presentation]. Proceedings of the TESOL Arabia conference, United Arab Emirates University.
Brown, J. D. (2002e). Statistics corner: Questions and answers about language testing statistics: The Cronbach alpha reliability estimate. *JALT Testing & Evaluation SIG Newsletter, 6*(1), 17-19.
Brown, J. D. (2002f).Do Cloze Tests Work? Or, is it just an illusion? *Second Language Studies, 21*(1), 79-125.

Brown, J. D. (2003a). Statistics corner: Questions and answers about language testing statistics: Norm-referenced item analysis: item facility and item discrimination. *Shiken: The JALT Testing & Evaluation SIG Newsletter, 7*(2), 16-19.

Brown, J. D. (2003b). Statistics corner: Questions and answers about language testing statistics: Criterion-referenced item analysis: The difference index and B. *Shiken: The JALT Testing & Evaluation SIG Newsletter, 7*(3),18– 24.

Brown, J. D. (2003c). The coefficient of determination. *Shiken: The JALT Testing & Evaluation SIG Newsletter, 7*(1), 14-16.

Brown, J. D. (2003d). The importance of criterion-referenced testing in East Asian EFL contexts. *ALAK News: A newsletter for the Applied Linguistics Association of Korea, 1,* 7-11.

Brown, J. D. (2003e). Creating a complete language-testing program. In C. A. Coombe & N. J. Hubley (Eds.), *Assessment practices* (pp.9-23). TESOL.

Brown, J. D. (2003f, May 11-12). *English language entrance examinations: A progress report* [Paper presentation]. Proceedings of the 2002 JALT Pan-SIG Conference, Kyoto, Japan.

Brown, J. D. (2003g). *The many facets of language curriculum development* [Paper presentation]. Proceedings of the 2003 5th CULI International Conference (pp.1-180. Chulalongkorn University.

Brown, J. D. (2003h). Voices in the field: An interview with J. D. Brown. *Shiken: JALT Testing & Evaluation SIG Newsletter, 7*(1), 10-13.

Brown, J. D. (2003i). The importance of criterion-referenced testing in East Asian EFL contexts. *ALAK Newsletter,* 7-11.

Brown, J. D. (2004a).Performance assessment: Existing literature and directions for research. *Second Language Studies, 22*(2), 91-139.

Brown, J. D. (2004b). Statistics corner: Questions and answers about language testing statistics: Yates correction factor. *Shiken: JALT Testing & Evaluation SIG Newsletter, 8*(1), 22-27.

Brown, J. D. (2004c). Statistics corner: Questions and answers about language testing statistics: Test-taker motivations. *Shiken: JALT Testing & Evaluation SIG Newsletter, 8*(2), 12-15.

Brown, J. D. (2004d). Research methods for applied linguistics: Scope, characteristics, and standards. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp.476-500). Blackwell Publishing.

Brown, J. D. (2004e). Grade inflation, standardized tests, and the case for on-campus language testing. In D. Douglas (Ed.), *English language testing in U.S. colleges and* universities (pp.37-56). NAFSA.

Brown, J. D. (2004f). What do we mean by *bias*, *Englishes*, *Englishes in testing*, and *English language proficiency*? *World Englishes, 23*(2), 317-319. https://doi.org/10.1111/j.0883-2919.2004.00355.x

Brown, J. D. (2004g). Resources on quantitative/statistical research for applied linguists. *Second Language Research, 20*(4), 372-393. https://doi.org/10.1191/0267658304sr245raff.ffhal-00572077

Brown, J. D. (2005a). Statistics corner: Questions and answers about language testing statistics: Generalizability and decision studies. *Shiken: JALT Testing & Evaluation SIG Newsletter, 9*(1), 12 – 16.

Brown, J. D. (2005b). *Testing in language programs: A comprehensive guide to English language assessment.* Prentice Hall Regents.

Brown, J. D. (2005c). Statistics corner: Questions and answers about language testing statistics: Characteristics of sound quantitative research. *Shiken, 9*(2), 31-33.

Brown, J. D. (2005d). Publishing without perishing. *Journal of the African Language Teachers Association, 6*(1), 1-16.

Brown, J. D. (2005e). Language testing policy: Administrators', teachers', and students' views. In S. H. Chan (Ed.), *ELT concerns in assessment* (pp.1-25). Sasbadi.

Brown, J. D. (2005f). Test review of the word meaning through listening test. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp.1159-1161). The Buros Institute of Mental Measurements.

Brown, J. D. (2006a). Teacher resources for teaching connected speech. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 27-47). National Foreign Language Resource Center.

Brown, J. D. (2006b). Introducing connected speech. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 1-15). National Foreign Language Resource Center.

Brown, J. D. (2006c). Testing reduced forms. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 247-264). National Foreign Language Resource Center.

Brown, J. D. (2006d). Second language studies: Curriculum development. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp.102—110). Elsevier.

Brown, J. D. (2006e). Statistics corner: Questions and answers about language testing statistics: Generalizability from second language research samples. *Shiken: JALT Testing & Evaluation SIG Newsletter, 10* (2), 24-27.

Brown, J. D. (2006f, May). *Promoting fluency in EFL classrooms* [Paper presentation]. Proceedings of the 2nd annual JALT Pan-SIG conference (pp. 1-12). Kyoto, Japan: Kyoto Institute of Technology.

Brown, J. D. (2006g, May 13-14). *Authentic communication: Whyzit importan'ta teach reduced forms* [Paper presentation]. Proceedings of the 5th annual JALT Pan-SIG conference (pp.13 - 24). Shizuoka, Japan: Tokai University College of Marine Science.

Brown, J. D. (2006h). Statistics corner: Questions and answers about language testing statistics: Resources available in language testing. *Shiken, 10*(1), 21-26.

Brown, J. D. (2007a). Statistics corner: Questions and answers about language testing statistics: Sample size and power. *Shiken: JALT testing & evaluation SIG newsletter, 11*(1), 31-35.

Brown, J. D. (2007b). Statistics corner: Questions and answers about language testing statistics: Sample size and statistical precision. *Shiken: JALT Testing & Evaluation SIG Newsletter, 11*(2), 21-24.

Brown, J. D. (2007c). Test review of Woodcock-Munoz language survey – revised. In R. A. Spies, B. S. Plake, K. F. Geisinger, & J. F. Carlson (Eds.), *The seventeenth mental measurements yearbook* (pp.872-874). The Buros Institute of Mental Measurements.

Brown, J. D. (2007d). Test review of English placement Test – revised. In R. A. Spies, B. S. Plake, K. F. Geisinger, & J. F. Carlson (Eds.), *The seventeenth mental measurements yearbook* (pp. 305-308). The Buros Institute of Mental Measurements.

Brown, J. D. (2007e). Multiple views of L1 writing score reliability. *Second Language Studies, 25*(2), 1-31.

Brown, J. D. (2008a). Statistics Corner. Questions and answers about language testing statistics: The Bonferroni adjustment. *Shiken: JALT Testing & Evaluation SIG Newsletter, 12*(1), 23-28.

Brown, J. D. (2008b). Statistics corner: Questions and answers about language testing statistics: Effect size and eta squared. *Shiken: JALT Testing & Evaluation SIG Newsletter, 12*(2), 36-41.

Brown J. D. (2008c). Raters, functions, item types and the dependability of L2 pragmatics tests. In E. Alcón Soler & A. Martínez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 224–248). Multilingual Matters.

Brown, J. D. (2008d). Testing-context analysis: Assessment is just another part of language curriculum development. *Language Assessment Quarterly, 5*(4), 275-312. https://doi.org/10.1080/15434300802457455

Brown, J. D. (2009a). Foreign and second language needs analysis. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp.269-293). Wiley-Blackwell.

Brown, J. D. (2009b). Statistics corner: Questions and answers about language testing statistics: Principal components analysis and exploratory factor analysis—Definitions, differences, and choices. *Shiken: JALT Testing & Evaluation SIG Newsletter, 13(1), 26-30.*

Brown, J. D. (2009c).Open-response items in questionnaires. In J. Heigham & R. A. Croker (Eds.), *Qualitative research in applied linguistics* (pp.200-219). Palgrave Macmillan. https://doi.org/10.1057/9780230239517_10

Brown, J. D. (2009d). Statistics corner: Questions and answers about language testing statistics: Principal components analysis and exploratory factor analysis—Choosing the right type of rotation in PCA and EFA. *Shiken: JALT Testing & Evaluation SIG Newsletter, 13*(3), 20-25.

Brown, J. D. (2009e).Statistics Corner. Questions and answers about language testing statistics: Choosing the right number of components or factors in PCA and EFA. *Shiken: JALT Testing & Evaluation SIG Newsletter, 13*(2), 19-23.

Brown. J. D. (2009f). Language curriculum development: Mistakes were made, problems faced, and lessons learned. *Second Language Studies, 28*(1), 85-105.

Brown, J. D. (2009g). Using a spreadsheet program to record, organize, analyze, and understand your classroom assessments. D. Lloyd., P. Davidson, & C. Coombe (Eds.), *The Fundamentals of language assessment: A practical guide for teachers* (2nd. Ed. pp.59-70). TESOL Arabia,

Brown, J. D. (2010a). Statistics Corner. Questions and answers about language testing statistics: How are PCA and EFA used in language test and questionnaire development? *Shiken: JALT Testing & Evaluation SIG Newsletter, 14*(2), 22-27.

Brown, J. D. (2010b). Statistics Corner. Questions and answers about language testing statistics: How are PCA and EFA used in language research? *Shiken: JALT Testing & Evaluation SIG Newsletter, 14*(1), 19-23.

Brown, J. D. (2010c). *Adventures in language testing: How I learned from my mistakes over 35 years* [Paper presentation]. Proceedings of the 2010 annual GETA international conference (pp. 18- 28). Gwangju, Korea, Global English Teachers' Association.

Brown, J. D. (2011a). Statistics corner: Questions and answers about language testing statistics: Likert items and scales of measurement? *Shiken: JALT Testing & Evaluation SIG Newsletter, 15*(1) 10-14.

Brown, J. D. (2011b).What do the L2 generalizability studies tell us? *Journal of Assessment and Evaluation in Education, 1,* 1-37.

Brown, J. D. (2011c).Quantitative research in second language studies. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp.190-206). Routledge.

Brown, J. D. (2011d). Statistics Corner. Questions and answers about language testing statistics: Confidence intervals, limits, and levels. *Shiken: JALT Testing & Evaluation SIG Newsletter, 15*(2), 23-27.

Brown, J. D. (2012a). *New ways in teaching connected speech.* TESOL International.

Brown, J. D. (Ed.). (2012b). *Developing, using, analyzing rubrics in language assessment with case studies in Asian and Pacific languages (Ed.).* National Foreign Language Resource Center.

Brown, J. D. (2012c). What teachers need to know about test analysis. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp.105-112). Cambridge University Press.

Brown, J. D. (2012d). Choosing the right type of assessment. In C. Coombe, P. Davidson & B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language* assessment. Cambridge University Press.

Brown, J. D. (2012e). Classical test theory. In G. Fulcher & F. Davidson (Eds.), *Routledge handbook of language testing*. (pp.303-315). Routledge.

Brown, J. D. (2012f). Commentary on the questionnaires chapter by Judy Ng. In R. Barnard, & A. Burns (Eds.), *Researching language teacher cognition and practice: International case studies* (pp.39-44). Multilingual Matters.

Brown, J. D. (2012g). EIL (English as an international language) curriculum development. In L. Alsagoff, S., L. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language* (pp. 147-167). Routledge. https://doi.org/10.4324/9780203819159

Brown, J. D. (2012h). What do distributions, assumptions, significance vs. meaningfulness, multiple statistical tests, causality, and null results have in common? *Shiken Research Bulletin, 16*(1), 28-33.

Brown, J. D. (2012i). How do we calculate rater/coder agreement and Cohen's kappa? *Shiken Research Bulletin, 16*(2), 30-36.

Brown, J. D. (2012j). Writing up a replication report. In G. Porte (Ed.), *Replication research in applied linguistics* (pp.173-197). Cambridge University Press.

Brown, J. D. (2013a). My twenty-five years of cloze testing research: So what? *International Journal of Language Studies, 7*(1), 1-32.

Brown, J. D. (2013b). Statistics corner: Questions and answers about language testing statistics: Solutions to problems teachers have with classroom testing. *Shiken: Research Bulletin,* 27, 27-33.

Brown, J. D. (2013c).Assessment and evaluation: Quantitative methods. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd.

Brown, J. D. (2013d). Research on computers in language testing: Past, present and future. In M. Thomas, H. Reinders, & Warschauer (Eds.), *Contemporary Computer-Assisted Language Learning* (pp. 73-95). Bloomsbury Academic.

Brown, J. D. (2013e).Cut scores on language tests. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd.

Brown, J. D. (2013f). Inference. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0534

Brown, J. D. (2013g). Sampling: Quantitative methods. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal1033

Brown, J. D. (2013h). Chi-square and related statistics for 2× 2 contingency tables. *Shiken Research Bulletin, 17*(10), 33-40.

Brown, J. D. (2013i). Teaching Statistics in Language Testing Courses. *Language Assessment Quarterly, 10* (3), 351-369. https://doi.org/10.1080/15434303.2013.769554

Brown, G. D. (2013j). Score Dependability and Decision Consistency. A.J. Kunnan (Ed.), *The companion to language assessment* (pp. 1182-1206). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118411360.wbcla085

Brown, J. D. (Ed.) (2013k). *New ways of classroom assessment, revised*. Alexandria.

Brown, G. D. (2014a).Classical theory reliability. A.J. Kunnan (Ed.), *The companion to language assessment* (pp.1-17). John Wiley & Sons, Inc.

Brown, J. D. (2014b). Differences in how norm-referenced and criterion-referenced tests are developed and validated. *Shiken, 18*(1), 29-33.

Brown, J. D. (2014c). The future of world Englishes in language testing. *Language Assessment Quarterly, 11*(1), 5-26. https://doi.org/10.1080/15434303.2013.869817

Brown, J. D. (2014d). *Mixed methods research for TESOL*. Edinburgh University Press.

Brown, J. D. (2014e). The Importance of Rubrics in Language Assessment. *국제한국어교육학회/국제학술발표논문집*, 3-13.

Brown, J. D. (2015a). Why bother learning advanced quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp.9-20). Routledge.

Brown, J. D. (2015b).Characteristics of sound quantitative research. *Shaken, 19*(2), 24-28.

Brown, J. D. (2015c). Deciding upon a research methodology. In J. d. Brown & C. Coombe (Eds.), *The Cambridge guide to research in language teaching and learning intrinsic eBook*. Cambridge University Press.

Brown, J. D. (2015d). *Le Fetuao Samoan language center assessment project*. University of Hawaii.

Brown, J. D. (2016a). Professional reflection: Forty years in applied linguistics. *International Journal of Language Studies, 10*(1), 1-14.

Brown, J. D. (2016b). Assessment in ELT: Theoretical options and sound pedagogical choices. W.A. Renandya & H.P. Widodo (Eds.), *English Language Teaching Today, English Language Education* (pp.67-82). Springer International Publishing. https://doi.org/10.1007/978-3-319-38834-2_6

Brown, J. D. (2016c). *Introducing needs analysis and English for specific purposes.* Routledge.

Brown, J. D. (2016d). Characteristics of sound mixed methods research. *Shiken, 20*(1), 21-23.

Brown, J. D. (2016e). Language testing and technology. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp.167-185). Routledge. https://doi.org/10.4324/9781315657899

Brown, J. D. (2016f). Consistency of measurement categories and subcategories. *Shiken, 20*(2), 50-52.

Brown, J. D. (2017a). Forty years of doing second language testing, curriculum, and research: So what? *Language Teaching, 50*(2), 276-289. https://doi.org/10.1017/S0261444816000471

Brown, J. D. (2017b).Developing and using rubrics: Analytic or holistic. *Shiken, 21*(2), 20-26.

Brown, J. D. (2017c). Consistency in research design: categories and subcategories. *Shiken, 21*(1), 23-28.

Brown, J. D. (2018a). Statistics corner: Questions and answers about language testing statistics: Calculating reliability of dictation tests: Does K-R21 work? *Shiken 22*(2), 14-19.

Brown, J. D. (2018b). Standardized and proficiency testing. In J. Liontas (Ed), *The TESOL encyclopedia of English language teaching* (pp.1-8). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0832

Brown, J. D. (2018c). Assessing Pragmatic Competence. In J. Liontas (Ed), *The TESOL encyclopedia of English language teaching* (1-7). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0384

Brown, J. D. (2018d). Cloze tests. In J. Liontas (Ed.), *The TESOL encyclopedia of English language teaching.* John Wiley & Sons, Inc.

Brown, J. D. (2019a).Assessment feedback. *Journal of Asia TEFL, 16*(1), 334-344.http://dx.doi.org/10.18823/asiatefl.2019.16.1.22.334

Brown, J. D. (2019b). Statistics corner: Questions and answers about language testing statistics: What is assessment feedback and where can I find out more about it? *Shiken, 23*(1), 45-49.

Brown, J. D. (2019c). Global Englishes and the international standardized English language proficiency tests. In F. Fang & H. P. Widodo (Eds.), *Critical Perspectives on Global Englishes in Asia: Language Policy, Curriculum, Pedagogy and Assessment* (pp.64-83). Multilingual Matters. https://doi.org/10.21832/9781788924108

Brown, J. D. (2019d). Overall English proficiency (whatever that is). *Shiken, 23*(2), 43-47.

Brown, J. D. (2020).World Englishes and international standardized English proficiency tests. In C. L. Nelson, Proshina, Z.G., & D. R. Davis (Ed.), *The handbook of world Englishes* (pp.703-724). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119147282.ch39

Brown, J. D. (2021a). Problems caused by ignoring descriptive statistics in language testing. In B. Lanteigne, C. Coombe, & J. D. Brown (Eds.), *Challenges in language testing around the world: Insights for language test users* (pp.15-24). Springer Singapore.

Brown, J. D. (2021b).English language proficiency: What is it? and Where do learners fit into it? In H. Mohebbi & C. Coombe (Eds.), *Research questions in language education and applied linguistics: A Reference Guide* (PP.323-327). Springer. https://doi.org/10.1007/978-3-030-79143-8_57

Brown, J. D. (2022a). English for specific purposes: Classroom needs analysis. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp.68-82). Routledge. https://doi.org/10.4324/9781003106609

Brown, J. D. (2022b). Why should we seriously consider teaching Chinese for specific purposes? In H. Wang., & C.U. Grosse (Eds.), *Chinese for business and professionals in the workplace* (pp.25-35). Routledge.

Brown, J. (2023). James Dean Brown's essential bookshelf: Connected speech. *Language Teaching*, 1-13. https://doi.org/10.1017/S0261444822000441

Brown, J. D., & Ahn, R. C. (2011). Variables that affect the dependability of L2 pragmatics tests. *Journal of Pragmatics, 43*(1), 198-217. https://doi.org/10.1016/j.pragma.2010.07.026

Brown, J. D., & Alaimaleata, E.T. (2015). The Le Fetuao Samoan language center oral proficiency interview. *Second Language Studies, 32*(2), 1-16.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning, 34*(4), 21–42. https://doi.org/10.1111/j.1467-1770. 1984.tb00350.x

Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing, 25*(3), 349–383. https://doi.org/10.1177/0265532208090157

Brown, J. D., & Coombe, C. (2015). The Cambridge guide to research in language teaching and learning intrinsic eBook. The Cambridge University Press.

Brown, J. D., Cunha, M. I. A., Frota, S.F.N., & Ferreira, A.B.F. (2001). The development and validation of a Portuguese version of the motivated strategies for learning questionnaire. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp.257-279). University of Hawaii.

Brown, J. D., & Crowther, D. (2022). *Shaping learners' pronunciation: Teaching the connected speech of North American English.* Routledge.

Brown, J. D. & Grüter, T. (2022). The same cloze for all occasions? Using the Brown (1980) cloze test for measuring proficiency in SLA research. *International Review of Applied Linguistics in Language Teaching*, *60*(3), 599-624. https://doi.org/10.1515/iral-2019-0026

Brown, J. D., & Hachima, A. (2005). Test review of the Arabic Proficiency Test. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 55-57). Hardbound.

Brown, J. D., & Hilferty, A. (1982). The effectiveness of teaching reduced forms of listening comprehension. *RELC Journal, 17*(2), 59-70. https://doi.org/10.1177/003368828601700204

Brown, J. D., & Hilferty, A. G. (1998). Reduced-forms dictations. In J.D. Brown (Ed.), *New ways of classroom assessment. Washington. TESOL*.

Brown, J. D., & Hilferty, A. (2006). The effectiveness of teaching reduced forms for listening comprehension. In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 51-58). National Foreign Language Resource Center.

Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communication*, *8*(4), 533–556.https://doi.org/10.1177/0741088391008004005

Brown, J. D., Hsu, W.L., & Harsch, K. (2017). 2017 ELIPT writing test revision project. *Second Language Studies, 35*(2), 1-30.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4), 653-675. https://doi.org/10.2307/3587999

Brown, J. D., & Hudson, T. (1999). Comments on James D. Brown and Thom Hudson's "The alternatives in language assessment". The authors respond. *TESOL Quarterly*, *33*(4),734- 735. https://doi.org/10.2307/3587884

Brown, J. D., Hudson, T., & Clark, M. (2004). Issues in placement survey. *National Foreign Language Resource Center.* Retrieved from http://nflrc.hawaii.edu/ NetWorks/NW41.pdf

Brown, J. D., Hudson, T., & Kim, Y. (1999). *Developing Korean performance assessment.* University of Hawaii.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing.* Cambridge University Press.

Brown, J. D., Hudson, T., Norris, J. M., & Bonk, W. J. (2000). Performance assessment of ESL and EFL students. *University of Hawaii Working Papers in ESL, 18*(2), 99-139.

Brown, J. D., & Kondo-Brown, K. (Eds.). (2006). *Perspectives on teaching connected speech to second language learners.* National Foreign Language Resource Center.

Brown,J. D., Janssen, G.,Trace, J., & Kozhevnikova, L. (2012). A preliminary study of cloze procedure as a tool for estimating English readability for Russian students. *Second Language Studies, 31*(1), 1-22.

Brown, J. D., Janssen, G., Trace, J., & Kozhevnikova, A. (2019). Using cloze passages to estimate readability for Russian university students: Preliminary study. *Focus on Language Education and Research, 1*(1), 16-27.

Brown, J. D., Davis, J.M., Takahashi, C., & Nakamura, K. (2012). *Upper-level EIKEN examinations: Linking, validating, and predicting TOEFL® IBT scores at advanced levels of the Eiken tests.* Eiken Foundation of Japan.

Brown, J. D., & Pennington, M. C. (1991). Developing effective evaluation systems for language programs. In M.C. Pennington (Ed.), *Building better language programs: Perspectives on evaluation in ESL.* National Association for Foreign Student Affairs.

Brown, J. D., Ramos,T., Cook, H. G., & Lockhart, C. (1991, April 9-12,). *Southeast Asian languages proficiency examinations* [Paper presentation]. Proceedings of the Regional Language Centre Seminar on Language Testing and Language Program Evaluation.

Brown, J. D., Robson, G., & Rosenkjar, P. R. (1996). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. *University of Hawaii working papers in ESL, 15*(1), 33-72.

Brown, J. D., Robson, G., & Rosenkjar, P. R. (2001). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp.361-398). University of Hawaii.

Brown, J. D., & Rodgers, T. S. (2003). *Doing second language research.* Oxford University Pres.

Brown, J. D., & Ross, J.A. (1993- April). *Decision dependability of subtests, tests, and the overall TOEFL test battery* [Paper presentation]. Proceedings of the annual meeting of the language testing research colloquium, Cambridge, England.

Brown, J. D., & Salmani Nodoushan, M. A. (2015). Language testing: The state of the art (An online interview with James Dean Brown). *International Journal of Language Studies, 9*(4), 133-143.

Brown, J. D., & Sato, C.J. (1990).Three approaches to task-based syllabus design. *University of Hawaii Working Papers in ESL, 9*.

Brown, J. D., & Trace, J. (2016). Fifteen ways to improve classroom assessment. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 490-505). Routledge.

Brown, J. D., & Trace, J. (2018). Connected-speech dictations for testing listening. In N. Spada, N. V. Deusen-Scholl, & L.Gurzynski-Weiss (Eds.), *Language learning and language teaching* (Vol. 50, pp. 45-63). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.50.04bro

Brown, J.D., Trace, J., Janssen, G., & Kozhevnikova, L. (2016). How well do cloze items work and why? In C. Gitsaki & C. Coombe (Eds.), *Current issues in language evaluation, assessment and testing: Research and practice* (pp.2-39). Cambridge Scholars Publishing.

Brown, J.D., & Wolfe-Quintero, K. (1997). Teacher portfolios for evaluation: A great idea or a waste of time. *Language Teachers, 21*(1), 28-30.

Brown, J. D., Yamashiro, A. D., & Ogane, E. (1999). Tailoring cloze: Three ways to improve cloze tests. *University of Hawaii Working Papers in ESL, 17*(2), 107-131.

Brown, J. D., Yamashiro, A. D., & Ogane, E. (2001). The emperor's new cloze: strategies for revising cloze tests. In T. Hudson., & J.D. Brown. *A focus on language test development: Expanding the language proficiency construct across a variety of tests (Vol.21)* (pp.143-161). University of Hawaii Press.

Brown, J.D., & Yamashita, S. O. (1995a). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal, 17*(1).

Brown, J. D., & Yamashita, S. O. (Eds.). (1995b). *Language testing in Japan.* The Japan Association for Language Teaching.

Brown, J. D., & Yamashita, S. O. (1995c). The authors respond to O'Sullivan's letter to JALT Journal: Out of criticism comes knowledge. *JALT Journal, 17*(2), 257-260.

Brown, J.D. & Yamashita, S.O. (1995d). English language entrance examinations at Japanese universities: 1993 and 1994. In J.D. Brown & S.O. Yamashita (Eds.) *Language testing in Japan* (pp.86-100). Japanese Association for Language Teaching.

Brown, J. D., Yongpei, C., & Yinglong, W. (1984). An evaluation of native-speaker self-access reading materials in an EFL setting. *RELC Journal, 15*(1), 75–84. https://doi.org/10.1177/003368828401500105

Bruton, A. (1999). Comments on James D. Brown and Thom Hudson's "The alternatives in language assessment": A reader reacts. *TESOL Quarterly, (33),* 4,729-734. https://doi.org/10.2307/3587884

Fulcher, G., Panahi, A., & Mohebbi, H. (2022). Glenn Fulcher's thirty-five years of contribution to language testing and assessment: A Systematic review. *Language Teaching Research Quarterly, 29*(2), 20-65. https://doi.org/10.32038/ltrq.2022.29.03

Housman, A., Dameg, K., Kobashigawa, M., & Brown, J.D. (2011). Report on the Hawaiian oral language assessment. *Second Language Studies, 29*(2), 1-59.

Hudson, T., & Brown, J. D. (Eds.). (2001). *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (Vol. 21). University of Hawaii Press.

Hudson, T., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics (Vol. 2).* Second Language Teaching and Curriculum Center.

Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics.* Second Language Teaching and Curriculum Center.

Iwai, T., Kondo, K., Lim, D. S. J., Ray, G. E., Shimizu, H., & Brown, J. D. (1999). *Japanese language needs analysis.* University of Hawaii.

Kachru, B. B. (1986). The power and politics of English. *World Englishes 5,* 121−140.

Kondo-Brown, K., & Brown, J. D. (2017). *Teaching Chinese, Japanese, and Korean heritage language students: Curriculum needs, materials, and assessment.* Routledge. https://doi.org/10.4324/9781315087443

Kondo-Brown, K., & Brown, J. D. (2000). *Japanese placement tests at the University of Hawai'i: Applying item response theory.* University of Hawaii press.

Kondo-Brown, K., Brown, J. D., & Tominaga, W. (2013). *Practical assessment tools for college Japanese.* National Foreign Language Resource Center.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. Psychmetrika, 2, 151-160. http://dx.doi.org/10.1007/BF02288391

Lanteigne, B., Coombe, C., & Brown, J. D. (Eds.) (2021a). *Challenges in language testing around the world: Insights for language test users*. Springer Singapore.

Lanteigne, B., Coombe, C., & Brown, J. D. (2021b). Reflecting on challenges in language testing around the world. In B. Lanteigne, C. Coombe, & J. D. Brown (Eds.), *Challenges in language testing around the world: Insights for language test users* (pp.549-553). Springer Singapore.

Mager, R. F. (1962). *Preparing educational objectives.* Fearon.

McKay, S., & Brown, J. D. (2015). *Teaching and assessment in EIL classrooms*. Routledge.

McKay, S. L., & Brown, J. D. (2016). *Teaching and assessing EIL in local contexts around the world*. Routledge

Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002a). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing, 19*(4), 395–418. https://doi.org/10.1191/0265532202lt237oa

Norris, J., Brown, J. D., Hudson,T. & Bonk, W.J. (2002b). *An investigation of second language task-based performance assessments.* University of Hawaii.

Norris, J. M., Brown J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments.* University of Hawaii.

Oliveira, L.P.D., Miller, I.K.D., Martins, M.D.A.P., Cunha, M.I.A., Brown, J. D. (1997). Recharging the battery: Placement tests for ESP students. *Alfa, São Paulo, 41,* 133-158.

Oller, J. W. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning, 23*(1), 105–118. https://doi.org/10.1111/j.1467-1770.1973.tb00100.x

Oller, J. W. (1979). *Language tests at school: a pragmatic approach.* Longman.

Oller, J. W. & Conrad, C. A. (1971). Cloze technique and ESL proficiency. *Language Learning, 21*(2), pp. 183-194. https://doi.org/10.1111/j.1467-1770.1971.tb00057.x

Palacio,M., Gaviria,S., & Brown, J.D. (2016). Aligning English language testing with curriculum. *Profile Issues in Teachers Professional Development, 18*(2), 63-77.

Pennington, M.C., & Brown, J.D. (1991).Unifying curriculum process and curriculum outcomes: The key to excellence in language education. In M.C. Pennington (Ed.), *Building better English language programs: Perspectives on evaluation in ESL* (pp. 57-74). National Association for Foreign Student Affairs.

Popham, W. J. (1975). *Educational evaluation* (1st ed.). Prentice-Hall.

Popham, W. J. (1978). *Criterion-referenced measurement.* Prentice-Hall.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement 6*, 1-9. https://doi.org/10.1111/j.1745-3984.1969.tb00654.x

Purpura, J. E., Brown, J. D., & Schoonen, R. (2015). Improving the Validity of Quantitative Measures in Applied Linguistics Research. *Language Learning, 65* (S1), 37-75. https://doi.org/10.1111/lang.12112

Trace, J., Hudson, T., & Brown, J. D. (2015). *Developing courses in language for specific purposes.* National Foreign Language Resource Center.

Trace, J., Brown, J. D., Janssen, G., & Kozhevnikova, L. (2017). Determining cloze item difficulty from item and passage characteristics across different learner backgrounds. *Language Testing, 34*(2), 151–174. https://doi.org/10.1177/0265532215623581

Trace, J., Brown, J. d., & Rodriguez, J. (2017a). How do stakeholder groups' views vary on technology in language learning? *Computer Assisted Language Instruction Consortium, 35*(2), 142–161. https://doi.org/10.1558/cj.32211

Trace, J., Brown, J. D., & Rodriguez, J. (2017b). Databank on stakeholder views of technology in language learning tools. *Second Language Studies, 36*(1), 53-73.

Wolfe-Quintero, K., & Brown, J. D. (1998). Teacher portfolios. *TESOL Journal, 7*(6), 24-27.

Youn, S. J., & Brown, J. D. (2013). Item difficulty and heritage language learner status in pragmatic tests for Korean as a foreign language. In S. J. Ross., & G. Kasper (Eds.), *Assessing second language pragmatics* (pp.98-123). Palgrave Macmillan. https://doi.org/10.1057/9781137003522_4

Appendix 1

1. Bailey, K.M, & Brown, J. D. (1996). Language Testing
2. Brown, J. D. (1980). Cloze Test
3. Brown, J. D. (1983a).Cloze Test
4. Brown, J.D. (1983b). Cloze Test
5. Brown, J.D. (1983c). NRTs
6. Brown, J. D. (1984a). Cloze Test
7. Brown, J.D. (1984b). CRTs
8. Brown, J. D. (1988a). Cloze Test
9. Brown, J. D. (1988b). Reading
10. Brown, J. D. (1988c). Research
11. Brown, J. D. (1988d). Cloze Test
12. Brown, J. D. (1989a). Language Program
13. Brown, J. D. (1989c). Cloze Test
14. Brown, J. D. (1989d). Writing
15. Brown, J. D. (1989e). Placement Test
16. Brown, J. D. (1989f). CRTs
17. Brown, J. D. (1989h). Curriculum
18. Brown, J. D. (1989i). Placement Test
19. Brown, J. D. (1990a). Placement Test
20. Brown, J. D. (1990b). Research
21. Brown, J. D. (1990c). CRTs
22. Brown, J. D. (1990d). Language Programs
23. Brown, J. D. (1990e). Placement
24. Brown, J.D. (1991a). Language Testing
25. Brown, J. D. (1991b). Writing
26. Brown, J. D. (1991c). Statistics
27. Brown, J. D. (1991d). Language Testing
28. Brown, J. D. (1992a). Computers
29. Brown, J. D. (1992b). Language Testing
30. Brown, J.D. (1992c).Technology
31. Brown, J. D. (1992d). Research
32. Brown, J. (1992e). TESOL
33. Brown, J. D. (1992f). TESOL
34. Brown, J.D. (1992g, March 1). Cloze Test
35. Brown, J. D. (1992h). Statistics
36. Brown, J. D. (1993a). Curriculum
37. Brown, J. D. (1993b). Language Testing
38. Brown, J. D. (1993c). CRTs
39. Brown, J. D. (1993d). Cloze Test
40. Brown, J.D., (1993e). CRTs
41. Brown, J. D. (1995a). Curriculum
42. Brown, J. D. (1995b). CRTs
43. Brown, J. D. (1995c). NRTs
44. Brown, J. D. (1995d). Entrance Exam
45. Brown, J.D. (1995e). Language Testing
46. Brown, J.D. (1995f). Language Program
47. Brown, J. D. (1995g). Curriculum
48. Brown, J.D. (1999g). Assessment
49. Brown, J. D. (1995h). Entrance Exam
50. Brown, J. D. (1995i). TESOL
51. Brown, J.D. (1995j). Norm
52. Brown, J. D. (1996a). Language Testing
53. Brown, J. D. (1996b). Entrance Exam
54. Brown, J. D. (1996c). Fluency
55. Brown, J. D. (1996d).Entrance Exam
56. Brown, J. D. (1997a). Research
57. Brown, J. D. (1997b). Reliability
58. Brown, J.D. (1997c). Language Programs
59. Brown, J. D. (1997d). Washback

60. Brown, J. D. (1997e). Statistics
61. Brown, J. D. (1997f). Washback
62. Brown, J. D. (1997g). Washback
63. Brown, J. D. (1997h). Computer
64. Brown, J. D. (1998a). Assessment
65. Brown, J. D. (1998b). Pragmatics
66. Brown, J. D. (1998c). Statistics
67. Brown, J. D. (1999a). Language Testing
68. Brown, J.D. (1999b). Assessment
69. Brown, J. D. (1999c). Statistics
70. Brown, J. D. (2000a). Washback.
71. Brown, J. D. (2000b). Questionnaire
72. Brown, J. D. (2000c). Validity
73. Brown, J. D. (2000d). Pragmatics
74. Brown, J. D. (2000e). Curriculum
75. Brown, J. D. (2000f). TESOL
76. Brown, J. D. (2001a). Pragmatics
77. Brown, J. D. (2001c). Statistics
78. Brown, J. D. (2001d). Reliability
79. Brown, J. D. (2001e). Language Programs
80. Brown, J. D. (2001f). Statistics
81. Brown, J. D. (2001g). Pragmatics.
82. Brown, J. D. (2001h). Language Testing
83. Brown, J. D. (2001i). CRTs
84. Brown, J. D. (2002a). Task-Based
85. Brown, J. D. (2002b). Washback
86. Brown, J. D. (2002c). Statistics.
87. Brown, J. D. (2002d). Language Testing
88. Brown, J. D. (2002e). Reliability
89. Brown, J. D. (2002f). Cloze Test
90. Brown, J. D. (2003a). NRTs
91. Brown, J. D. (2003b). CRTs
92. Brown, J. D. (2003c). Statistics
93. Brown, J. D. (2003d). CRTs
94. Brown, J. D. (2003e). Language Testing
95. Brown, J. D. (2003f, May 11-12). Entrance Exam
96. Brown, J. D. (2003g). Curriculum
97. Brown, J. D. (2003h). Language Testing
98. Brown, J. D. (2003i). CRTs
99. Brown, J. D. (2004a). Performance
100. Brown, J. D. (2004b). Statistics
101. Brown, J. D. (2004c). TESOL
102. Brown, J. D. (2004d). Applied Linguistics
103. Brown, J. D. (2004e). Language Testing
104. Brown, J. D. (2004f). TESOL
105. Brown, J. D. (2004g). Research
106. Brown, J. D. (2005a). Generalizability
107. Brown, J. D. (2005b). Language Testing
108. Brown, J. D. (2005c). Research
109. Brown, J. D. (2005d). Research
110. Brown, J. D. (2005e). Language Testing
111. Brown, J. D. (2005f). Listening
112. Brown, J. D. (2006a). Connected Speech
113. Brown, J. D. (2006b). Connected Speech
114. Brown, J. D. (2006c). Connected Speech
115. Brown, J. D. (2006d). Curriculum
116. Brown, J. D. (2006e). Generalizability
117. Brown, J. D. (2006f). Fluency
118. Brown, J. D. (2006g). Connected Speech
119. Brown, J. D. (2006h). Language Testing

120. Brown, J. D. (2007a). Language Testing
121. Brown, J. D. (2007b). Statistics
122. Brown, J. D. (2007c). Language Testing
123. Brown, J. D. (2007d). Placement Test
124. Brown, J. D. (2007e). Writing
125. Brown, J. D. (2008a). Statistics
126. Brown, J. D. (2008b). Statistics
127. Brown J. D. (2008c). Pragmatics
128. Brown, J. D. (2008d). Curriculum
129. Brown, J. D. (2009a). TESOL
130. Brown, J. D. (2009b). Factor Analysis
131. Brown, J. D. (2009c). Questionnaire
132. Brown, J. D. (2009d). Factor Analysis
133. Brown, J. D. (2009e). Factor Analysis
134. Brown. J. D. (2009f). Curriculum
135. Brown, J. D. (2009g). Statistics
136. Brown, J. D. (2010a). Factor Analysis
137. Brown, J. D. (2010b). Factor Analysis
138. Brown, J. D. (2010c). Language Testing
139. Brown, J. D. (2011a). Statistics
140. Brown, J. D. (2011b). Generalizability
141. Brown, J. D. (2011c). Research
142. Brown, J. D. (2011d). Statistics
143. Brown, J. D. (2012a). Connected Speech
144. Brown, J. D. (Ed.). (2012b). Rubrics
145. Brown, J. D. (2012c). Language Testing
146. Brown, J. D. (2012d). Assessment
147. Brown, J. D. (2012e). Classical Test Theory
148. Brown, J. D. (2012f). Questionnaires
149. Brown, J. D. (2012g). Curriculum
150. Brown, J. D. (2012h). Statistics
151. Brown, J. D. (2012i). Statistics
152. Brown, J. D. (2012j). Research
153. Brown, J. D. (2013a). Cloze Test
154. Brown, J. D. (2013b). Assessment
155. Brown, J. D. (2013c). Assessment
156. Brown, J. D. (2013d). Computer
157. Brown, J. D. (2013e). Language Testing
158. Brown, J. D. (2013f). Statistics
159. Brown, J. D. (2013g). Research
160. Brown, J. D. (2013h). Statistics
161. Brown, J. D. (2013i). Statistics
162. Brown, G. D. (2013j). Reliability
163. Brown, J. D. (Ed.) (2013k). Assessment
164. Brown, G. D. (2014a).Classical Test Theory
165. Brown, J. D. (2014b). NRTs
166. Brown, J. D. (2014c). TESOL
167. Brown, J. D. (2014d). Research
168. Brown, J. D. (2014e). Rubrics
169. Brown, J. D. (2015a). Research
170. Brown, J. D. (2015b). Research
171. Brown, J. D. (2015c). Research
172. Brown, J. D. (2015d). Assessment
173. Brown, J. D. (2016a). Professional Reflection
174. Brown, J. D. (2016b). Assessment
175. Brown, J. D. (2016c). ESP
176. Brown, J. D. (2016d). Research
177. Brown, J. D. (2016e). Language Testing
178. Brown, J. D. (2016f). Reliability
179. Brown, J. D. (2017a). Professional Reflection
180. Brown, J. D. (2017b). Rubrics

181. Brown, J. D. (2017c). Reliability
182. Brown, J. D. (2018a). Reliability
183. Brown, J. D. (2018b). Proficiency Test
184. Brown, J. D. (2018c). Pragmatics
185. Brown, J. D. (2018d). Cloze Test
186. Brown, J. D. (2019a). Feedback
187. Brown, J. D. (2019b). Feedback
188. Brown, J. D. (2019c). Standardized Test
189. Brown, J. D. (2019d). TESOL
190. Brown, J. D. (2020). Standardized test
191. Brown, J. D. (2021a). Statistics
192. Brown, J. D. (2021b).TESOL
193. Brown, J. D. (2022a). ESP
194. Brown, J. D. (2022b). ESP
195. Brown, J. (2023). Connected Speech
196. Brown, J. D., & Ahn, R. C. (2011). Pragmatics Tests
197. Brown, J. D., & Alaimaleata, E.T. (2015). Oral Proficiency
198. Brown, J. D., & Bailey, K. M. (1984). Writing
199. Brown, J. D., & Bailey, K. M. (2008). Language Testing
200. Brown, J. D., & Coombe, C. (2015). TESOL
201. Brown, J. D., Cunha, M. I. A., Frota, S.F.N., & Ferreira, A.B.F. (2001). TESOL
202. Brown, J. D., & Crowther, D. (2022). Connected Speech
203. Brown, J. D. & Grüter, T. (2022). Cloze Test
204. Brown, J. D., & Hachima, A. (2005). Proficiency Test
205. Brown, J. D., & Hilferty, A. (1982). Reduced Forms
206. Brown, J. D., & Hilferty, A. G. (1998). Reduced Forms
207. Brown, J. D., & Hilferty, A. (2006). Listening
208. Brown, J. D., Hilgers, T., & Marsella, J. (1991). Writing
209. Brown, J. D., Hsu, W.L., & Harsch, K. (2017). Writing
210. Brown, J. D., & Hudson, T. (1998). Assessment
211. Brown, J. D., & Hudson, T. (1999). Assessment
212. Brown, J. D., Hudson, T., & Clark, M. (2004). Placement Test
213. Brown, J. D., Hudson, T., & Kim, Y. (1999). Performance Test
214. Brown, J. D., & Hudson, T. (2002).CRTs
215. Brown, J. D., Hudson, T., Norris, J. M., & Bonk, W. (2000g). Performance Test
216. Brown,J. D., Hudson,T. Norris, J., & Bonk, W.J. (2002). Task-Based
217. Brown, J. D., & Kondo-Brown, K. (Eds.). (2006). Connected Speech
218. Brown,J. D., Janssen, G.,Trace, J., & Kozhevnikova, L.(2012). Cloze Test
219. Brown, J. D., Janssen, G. , Trace, J., & Kozhevnikova, A. (2019). Reliability
220. Brown, J. D. Davis, J.M., Takahashi, C., & Nakamura, K. (2012). Validity
221. Brown, J. D., & Pennington, M. C. (1991). Language Program
222. Brown,J.D., Ramos,T., Cook, H.G., & Lockhart, C. (1991). Proficiency Test
223. Brown, J. D., Robson, G., & Rosenkjar, P. R. (1996). TESOL
224. Brown, J. D., Robson, G., & Rosenkjar, P. R. (2001). TESOL
225. Brown, J. D., & Rodgers, T. S. (2003). Research
226. Brown, J. D., & Ross, J.A. (1993- April). Language Testing
227. Brown, J. D., & Salmani Nodoushan, M. A. (2015). Language Testing
228. Brown, J. D., & Sato, C.J. (1990). Task-Based Syllabus
229. Brown, J. D., & Trace, J. (2016). Assessment
230. Brown, J. D., & Trace, J. (2018). Connected Speech
231. Brown, J.D., Trace, J., Janssen, G., & Kozhevnikova, L. (2016). Cloze Test
232. Brown, J.D., & Wolfe-Quintero, K. (1997). Portfolio
233. Brown, J. D., Yamashiro, A. D., & Ogane, E. (1999). Cloze Test
234. Brown, J. D., Yamashiro, A. D., & Ogane, E. (2001). Cloze Test
235. Brown, J.D., & Yamashita, S. O. (1995a). Entrance Exam
236. Brown, J. D., & Yamashita, S. O. (Eds.). (1995b). Language Testing
237. Brown, J. D., & Yamashita, S. O. (1995c). Language Testing
238. Brown, J.D. & Yamashita, S.O. (1995d). Entrance Exam
239. Brown, J. D., Yongpei, C., & Yinglong, W. (1984). TESOL
240. Bruton, A. (1999). Assessment

241. Housman, A., Dameg, K., Kobashigawa, M., & Brown, J.D. (2011). Oral Proficiency
242. Hudson, T., & Brown, J. D. (Eds.). (2001). Language Testing
243. Hudson, T., Detmer, E., & Brown, J. D. (1992). Pragmatics
244. Hudson, T., Detmer,E., & Brown, J.D. (1995). Pragmatics
245. Iwai, T., Kondo, K., Lim, D.S. J., Ray, G. E., Shimizu, H., & Brown, J. D. (1999). Needs Analysis
246. Kondo-Brown, K., & Brown, J. D. (2017). Curriculum
247. Kondo-Brown, K., & Brown, J. D. (2000). Placement Tests
248. Kondo-Brown,K., Brown, J. D., & Tominaga, W. (2013). Assessment
249. Lanteigne, B., Coombe, C., & Brown, J. D. (Eds.) (2021a). Language Testing
250. Lanteigne, B., Coombe, C., & Brown, J. D. (2021b). Language Testing
251. McKay, S., & Brown, J. D. (2015). Language Testing
252. McKay, S. L., & Brown, J. D. (2016). Language Testing
253. Norris, J. M., Brown, J. D., Hudson, T. D., & Bonk, W. (2002). Task-Based
254. Norris, J. M., Brown J. D., Hudson, T., & Yoshioka, J. (1998). Performance Test
255. Oliveira, L.P.D., Miller, I.K.D., Martins, M.D.A.P., Cunha, M.I.A., Brown, J. D. (1997). Placement Test
256. Palacio,M., Gaviria,S., & Brown, J.D. (2016). Curriculum
257. Pennington, M.C., & Brown, J.D. (1991). Curriculum
258. Purpura, J. E., Brown, J. D., & Schoonen, R.(2015). Research
259. Trace, J., Hudson, T., & Brown, J. D. (2015). ESP
260. Trace, J., Brown, J. D., Janssen, G., & Kozhevnikova, L. (2017). Cloze Test
261. Trace, J., Brown, J. d., & Rodriguez, J. (2017a). Technology
262. Trace, J., Brown, J. D., & Rodriguez, J. (2017b). Technology
263. Wolfe-Quintero, K., & Brown, J. D. (1998). Portfolio
264. Youn, S. J., & Brown, J. D. (2013). Pragmatics