CONTENTS

Preface ............... ix

The Unspoken Element of Communicative Competence:
Evaluating Language Learners' Nonverbal Behavior
Nicholas O. Jungheim .......... 1

Using Pictures for Research in Pragmatics:
Eliciting Pragmatic Strategies by Picture Response Tests
Sayoko Yamashita ............. 37

Self-assessment Methods in Cross-cultural Pragmatics
Thom Hudson ............... 59

Measuring L2 Vocabulary
Yoshimitsu Kudo ............. 79

Testing ESL Learners' Knowledge of Collocations
William J. Bonk ............. 113

The Emperor's New Cloze: Strategies for Revising Cloze Tests
James Dean Brown, Amy D. Yamashiro, & Ethel Ogane .......... 143

Identifying Rating Criteria for Task-based EAP Assessment
John M. Norris ............. 163

Developing and Revising Criterion-referenced
Achievement Tests for a Textbook Series
James Dean Brown .......... 205

About the Contributors ........... 229
PREFACE

This volume originated out of discussions between the editors concerning two basic issues. The first issue is an interest over several years regarding non-traditional forms of second and foreign language assessment. The second issue came from a realization that testing projects are seldom reported in the literature if the project is not successful. Such studies become part of the bottom drawer file, never to see the light of day. Further, many projects tend to report only the final results without providing a view of the entire test development process, warts and all. With these two issues in mind, we contacted people who were involved in many different forms of language testing projects. Our goals were to present novel forms of language assessment and, to the extent possible, show the entire process of development. Also, we wanted to include language testing projects that demonstrated the types of problems that test developers have run into in the real world. Each of the papers in this volume represents a test development project that provides unique insights into the language test development process.

In The unspoken element of communicative competence: Evaluating language learners' nonverbal behavior, Nicholas O. Junghem offers an approach for assessing nonverbal communication. His argument is that although nonverbal communication is considered to be a part of communicative competence and evidence shows that it can have an influence on the rating of oral proficiency, language testers have basically ignored its role. The lack of a framework for assessing language learner's nonverbal behavior has impeded progress in this area. The purpose of his study is to present an assessment framework for nonverbal ability, describe the construction nonverbal ability scales in terms of a limited number of nonverbal behaviors, investigate the relationship between nonverbal ability and oral L2 proficiency, and examine the scales' validity and reliability. The results of this study indicate that language learners' nonverbal behavior can be accurately assessed for a limited number of behaviors within this framework.

In Using pictures for research in pragmatics: Eliciting pragmatic strategies with picture response tests, Sayoko Yamashita addresses issues surrounding the use of discourse completion tests (DCT) in second language pragmatic assessment. The first concern is that the realization of speech acts in written forms might result in differences from oral language production. Further, the traditional DCT format of a written description of a situation presents difficulties for beginner and intermediate learners. In order to get around some of these problems, Yamashita developed a set of pictures that frame the speech event and then examined their utility in cross-cultural pragmatics assessment.

Thom Hudson examines the role of self-assessment in the study of cross-cultural pragmatics in Self-assessment methods in cross-cultural pragmatics. There has been a great deal of interest in the promise of self-assessment as a means of assessment other than more traditional paper and pencil tests. Hudson examines the
relationships that exist between a rated assessment of each examinee on a task designed to elicit cross-cultural pragmatic performance and two types of self-assessment. The results do not show a strong relationship between the three measures, but this may be a result of low variance in the performance ratings.

In *Measuring L2 vocabulary*, Yoshimitsu Kudo adapts Nation’s (1983) test in order to assess the scalability of the levels presented in Nation’s list and to develop a test of vocabulary at a level appropriate for beginning level EFL learners. He found that it was necessary to avoid English loan words in Japanese in order to get a set of scalable levels. His results indicate that it is possible to find distinct levels of vocabulary, although individual items included in a defined level may not scale as well as others.

William Bonk reports on the development, administration, and analysis of a test of collocational knowledge for ESL learners of a wide range of proficiency levels. Through native speaker item validation and pilot testing, three subtests were developed and administered to 98 ESL learners of low-intermediate to advanced proficiency. Descriptive statistics and reliability estimates for the test administration are calculated, and the characteristics of the test items, subtests, and response modes were examined using traditional item analysis, item response theory, and generalizability theory methods. Two of the three subtests were found to perform well as norm-referenced measures of the construct, and areas for further testing and research were pinpointed. Observed collocational knowledge was found to correlate strongly ($r=.73$) with a measure of general ESL proficiency, while length of residence alone had negligible predictive power of collocations test performance. Exploratory factor analysis revealed that the collocations items tended to load on a different factor from general proficiency items, giving preliminary evidence of construct validity.

James D. Brown, Amy Yamashiro, and Ethel Ogane examine three approaches to the development of cloze tests: the hit-or-miss method, the modification method, and the tailored cloze method. They further examined the effects of using acceptable-answer versus exact-answer scoring. The results indicate that all three cloze tests, regardless of the method used for selection, needed to undergo modification in order to improve the fit of the score distribution to the particular audience of students.

John Norris descriptively explores the process involved in identifying the criteria used in evaluating examinee performances in *Identifying rating criteria for task-based EAP assessment*. The process involved in the negotiations between three creators of assessment criteria is described. He observed that the criteria negotiation session reflected a consensus orientation among the team members. Further, in general, the participants based their criteria for task accomplishment on task-centered criteria.

In *Developing and revising criterion-referenced achievement tests for a textbook series*, James Dean Brown describes the process of developing reliable and valid tests for a
set of published English language textbooks from the initial test rationale through piloting and revising the tests. He found that, from a criterion-referenced perspective, the tests were dependable in addition to being content, construct, and decision valid. The final forms of the tests appear to serve as appropriate achievement tests for the textbook series.