Assessing Language through Computer Technology

THE CAMBRIDGE LANGUAGE ASSESSMENT SERIES

Series editors: J. Charles Alderson and Lyle F. Bachman

In this series:

Assessing Grammar by James E. Purpura
Assessing Language for Specific Purposes by Dan Douglas
Assessing Listening by Gary Buck
Assessing Reading by J. Charles Alderson
Assessing Speaking by Sari Luoma
Assessing Vocabulary by John Read
Assessing Writing by Sara Cushing Weigle
Assessing Young Language Learners by Penny McKay
Statistical Analyses for Language Assessment by Lyle F. Bachman
Statistical Analyses for Language Assessment Workbook by Lyle F. Bachman and Antony J. Kunnan

Assessing Language through Computer Technology

Carol A. Chapelle
and
Dan Douglas

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

www.cambridge.org
Information on this title: www.cambridge.org/9780521549493

First published 2006

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication Data

Chapelle, Carol A., 1955-
Assessing language through computer technology / Carol A. Chapelle and Dan Douglas.
p. cm. – (The Cambridge language assessment series)
Includes bibliographical references and index.
ISBN-10: 0-521-54949-3 (alk. paper)
ISBN-10: 0-521-84021-X (alk. paper)
ISBN-13: 978-0-521-84021-7
ISBN-13: 978-0-521-54949-3
1. Language and languages – Ability testing – Data processing. I. Douglas, Dan. II. Title. III. Series.

P53.4.C394 2006
418.00285–dc22 2006002810

For Annie and Webster

Contents

Series editors' preface

Although many people think of computer technology as a recent development, computer technology has been used in language assessment for a long time. Large mainframe computers have been used since the 1960s for the analysis of test data and for the storage of items in databases, or item banks, as well as for producing reports of test results for test users. More recently with the advent of the personal computer, it is now common to use word-processing software for the creation and modification of test tasks, as well as for all of the tasks previously done on a mainframe computer. Perhaps the most striking change that computer and information technology has brought to language assessment, however, is the potential for delivering a wide variety of test tasks online anywhere in the world, and providing immediate feedback, 24 hours a day, seven days a week. This potential for expanding the kinds of tasks that we can deliver to test takers has been accompanied by ever-increasing capacity for scoring responses by computer. While selected-response items have been scored by optical scanners for half a century, recent advances in natural language processing and latent semantic analysis, along with improvements in scanning technology, have made it possible to score written responses to constructed response tasks – both open-ended and short-answer tasks and responses to composition prompts – by computer.

However, along with these potential advances come potential problems and concerns. To what extent might the use of multimedia in assessment tasks introduce features that actually diminish the validity of the interpretations and uses we make of assessment results? For example,

how does the use of video in a listening test affect test takers' performance? Does watching a video involve the same abilities as listening to a tape recording? If not, can we interpret scores from audio- and video-delivered tests in the same way? How do we know that scores of written essays produced by human raters can be interpreted in the same way as those produced by computers?

Thus, assessing language through computer technology is not without controversy, and debate is ongoing as to whether the advantages of computer-assisted language testing (CALT) outweigh the disadvantages. Critics argue that CALT constitutes a conservative element in test design, since test items are limited to those types which can be marked by machine. Proponents argue that CALT allows the incorporation of multimedia into tests, that the provision of immediate feedback presents significantly added value to the user, and that as technology develops, even the limitations of computer scoring will diminish as more intelligent scoring algorithms are developed. One frequently proclaimed advantage of CALT is the ability to tailor the test to the individual test taker, in computer-adaptive testing, by selecting the next item to which a test taker is exposed in the light of his or her response to the previous item. Thus if a test taker gets the item wrong, he or she will be presented with an easier item, whereas if the response is correct, the test taker will be presented with a more difficult item. This oft-proclaimed advantage has, however, serious costs associated with it. Firstly, test items need to be trialed on large numbers of test takers to ensure that the items are stable and accurately measure what they are intended to measure. Secondly, for high-stakes tests, very large numbers of pre-tested items are needed to refresh the item banks to ensure test security.

Such is an example of the discussions that surround language assessment through technology, and so this latest addition to the Cambridge Language Assessment Series is very timely. Carol Chapelle and Dan Douglas are experts in the field of language testing in general and in technology-based assessment in particular, having published widely, and having taught courses in these areas at universities and international institutes around the world. In addition, these two authors bring to this volume a wealth of combined knowledge about computer-assisted language learning and second language acquisition research. Furthermore, both have extensive experience in teacher training and in working with practitioners. Their combined experience has been brought to bear on a topic that can only increase in importance and impact. Therefore any person involved in language assessment, at whatever educational level,

will need to be familiar with the possibilities and limitations that they should consider when deciding whether to construct or use assessment procedures using computer technology.

The authors present a critical review of research that has sought to address controversial issues, such as whether computer-assisted language tests are equivalent to paper-and-pencil-tests, whether CALT can enhance test validity, what impact CALT might have, and they discuss these issues at length, from both theoretical and practical perspectives. Readers are also given a detailed account of test-authoring software and made aware of the advantages of such systems. Above all, Chapelle and Douglas discuss the ways in which CALT should be evaluated, and how traditional views of test validity need to be both taken into account and adjusted in light of the challenges presented by assessment using technology.

This book will be required reading for any test developer.

J. Charles Alderson
Lyle F. Bachman

Acknowledgments

We would like to extend our sincere thanks to Charles Alderson and Lyle Bachman for their guidance throughout our writing of this last book in the Cambridge Language Assessment Series. The complex topic addressed in this volume is one about which no single person has thorough expertise, and we have therefore benefited greatly from the assistance provided by Charles and Lyle. As authors of the final book in the series, we also take this opportunity to salute their foresight and ambition in proposing such a series and in seeing it through with such rigorous scholarly attention.

Our knowledge and interests in computer technology for assessing language abilities have developed over the past twenty-plus years as we have worked on projects with various people. Most notable of our collaborators are Volker Hegelheimer and Joan Jamieson, both of whom have taken us through ambitious technology projects with their management finesse, technical knowledge, and can-do spirit. Through work with them, we have learned a lot of what is reflected in this volume, which we hope will in turn teach others. We have also benefited through many years of rich interaction with the TOEFL program at Educational Testing Service and with Pearson Education.

We would like to express our appreciation to Mickey Bonin, who was the Cambridge editor in charge of the series when we began this project, and to Jane Walsh, who now oversees the project. We thank Marie Allan, Jacque French, and Clive Rumble for their professional and efficient help with editing and production.

Finally, we express our thanks to the following for permission to reprint

their copyrighted material: ACT Inc., Discovery School, Educational Testing Service, Half-baked, IBM Corporation, John Benjamins Publishing, *Journal of Technology, Learning and Assessment*, Ordinate Corporation, Respondus, WebCT, and *Studies in Language Learning*.